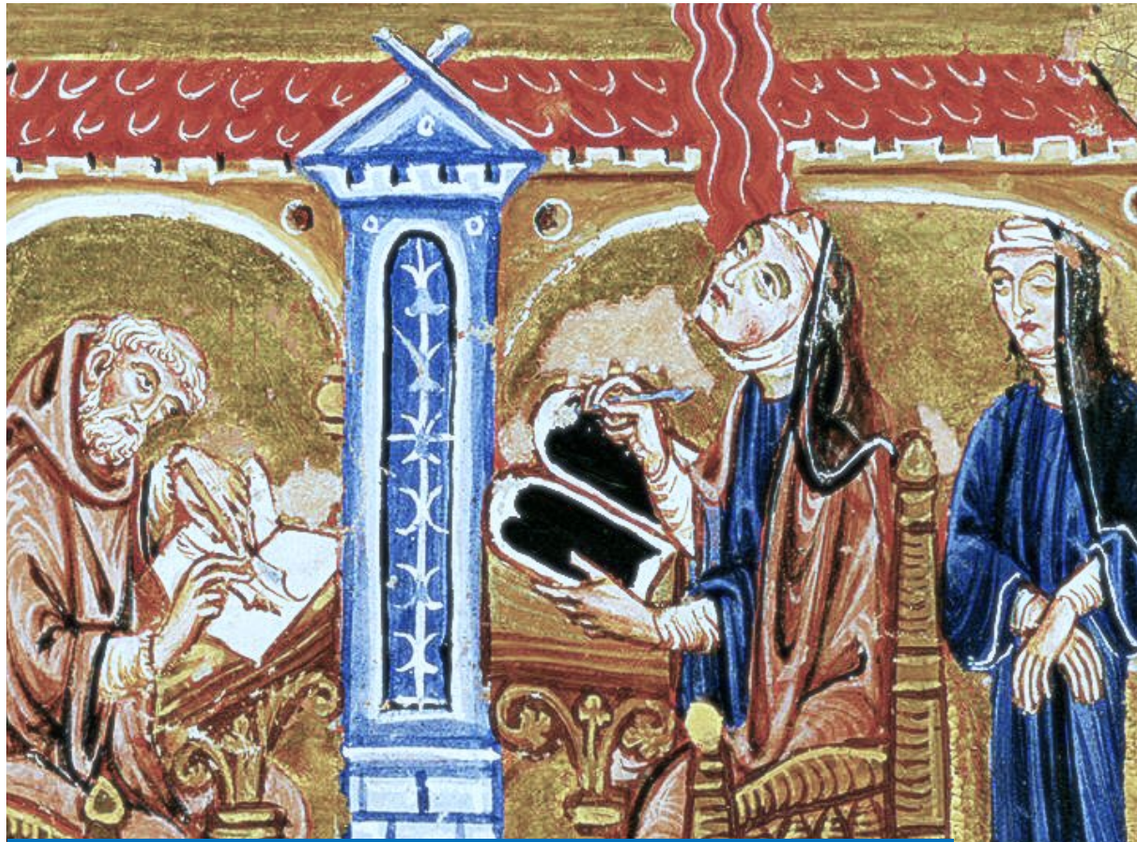
**Collaborative Authorship in Twelfth-Century Latin Literature**
A Stylometric Approach to Gender, Synergy and Authority

Jeroen De Gussem

FACULTEIT LETTEREN
EN WIJSBEGEERTE



# COLLABORATIVE AUTHORSHIP IN TWELFTH-CENTURY LATIN LITERATURE

A Stylometric Approach to Gender, Synergy and Authority

**Jeroen De Gussem**

Proefschrift voorgelegd tot het behalen van de interdisciplinaire graad van Doctor in de geschiedenis en letterkunde aan de Universiteit Gent en de graad van Doctor in de taal- en letterkunde aan de Universiteit Antwerpen

UNIVERSITEIT
GENT

Universiteit
Antwerpen

| | |
|---|---|
| Promotor: | Prof. dr. Jeroen Deploige<br>Vakgroep Geschiedenis, Universiteit Gent |
| Co-promotoren: | Prof. dr. Wim Verbaal<br>Vakgroep Letterkunde, Universiteit Gent |
| | Prof. dr. Mike Kestemont<br>Departement Letterkunde, Universiteit Antwerpen |
| Decaan: | Prof. dr. Gita Deneckere, Universiteit Gent |
| | Prof. dr. Luc Duerloo, Universiteit Antwerpen |
| Rector: | Prof. dr. Rik Van de Walle, Universiteit Gent |
| | Prof. dr. Herman Van Goethem, Universiteit Antwerpen |

Cover image: Hildegard of Bingen receiving inspiration from the Holy Ghost, and engraving her texts on a wax tablet by aid of a male and female secretary.

FACULTEIT LETTEREN
EN WIJSBEGEERTE

# Collaborative Authorship in Twelfth-Century Latin Literature

## A Stylometric Approach to Gender, Synergy and Authority

Jeroen De Gussem

Proefschrift voorgelegd tot het behalen van de interdisciplinaire graad van
Doctor in de Geschiedenis en Letterkunde aan de Universiteit Gent
en de graad van Doctor in de Taal- en Letterkunde aan de Universiteit Antwerpen, 2019

UNIVERSITEIT
GENT

Universiteit
Antwerpen

# Dedication

Deze thesis is in zekere mate ook zelf het resultaat van 'collaboratief auteurschap.' Daarmee bedoel ik niet dat iemand nu prompt de knop van de Ephorus plagiaatdetectiemachine moet gaan indrukken, maar wel dat ik dit werk nooit had kunnen schrijven zonder mij te beroepen op enkele *auctoritates* in mijn leven. Zonder het zelf misschien altijd te beseffen hebben deze mensen indirect mee geschreven aan de pagina's die dit boek beslaan.

In de loop van de afgelopen vier jaar kon ik staan op de schouders van reuzen Jeroen Deploige, Wim Verbaal en Mike Kestemont, mijn promotoren, die ik heb leren bewonderen niet enkel als de absolute top in hun respectievelijke vakgebieden, maar ook als mensen. Ik ben mijn naamgenoot en hoofdpromotor Jeroen Deploige, —of, moet ik zeggen, Hieronymus maior?—, erkentelijk voor het eindeloze enthousiasme waarmee hij al de vreemde curves die ik had gegenereerd analyseerde, steeds met de scherpe zin voor historische kritiek die al eens met een frons op zijn voorhoofd durft gebeiteld te staan. Jeroen liet me zien, met de vriendelijkheid en het geduld die hem typeren, dat de schoonheid van de twaalfde eeuw in de details schuilgaat, en het is zijn verdienste dat de letterkundige in mij inmiddels ook een beetje historicus werd. Wim Verbaal is bovenal mijn *magister* in het Latijn. Nog voor hij mijn promotor werd en mijn docent was, keek ik naar Wim op vanwege zijn hartstochtige passie voor het vak, de gedrevenheid en humor waarmee hij mijzelf en mijn medestudenten wist te begeesteren, en zijn talent om wat we voor vanzelfsprekend namen keer op keer te bevragen. Het maakt me nog steeds erg trots dat een intelligente man als hij in me geloofde wanneer ik solliciteerde voor dit project. Mike Kestemont, tenslotte, is een absolute duizendpoot en een bodemloos vat vol creativiteit: wie anders behalve hij weet dubstep te combineren met Hildegard van Bingen, of vegan barbecue met Bernardus van Clairvaux, en komt er nog mee weg ook? Er zijn, nu zie ik, te weinig wetenschappers die met de humor en het relativeringsvermogen van Mike door academia stappen, en toch zo doeltreffend datzelfde vermogen gebruiken om baanbrekend onderzoek te verrichten. Ik zal niet gauw vergeten dat aan het einde van mijn traject, wanneer het zwaarder begon te wegen, er hem tussendoor de technische feedback één zin ontviel die me misschien nog het meest heeft geholpen: "Maar gij neemt dat zo serieus allemaal!"

Dit project had niet kunnen slagen zonder de professionaliteit en collegialiteit van

talloze anderen. Zo ben ik ten eerste het Bijzonder Onderzoeksfonds erkentelijk, omwille van de financiering van het project waarop ik werd aangeworven, en het reiskrediet dat me verleend werd in 2017 om deel te kunnen nemen aan het DH congres in Montréal, een buitenkans om mijn expertise verder aan te scherpen. Ook het Fonds Wetenschappelijk Onderzoek en het Henri Pirenne Institute for Medieval Studies ben ik dankbaarheid verschuldigd omwille van hun genereuze co-financiering van het congres "The Medieval Literary Canon in the Digital Age" in 2018 dat me de kans heeft geboden enkele van de groten in het vakgebied uit te nodigen. Verder wil ik ook het "Corpus Christianorum Bibliotheek & Kenniscentrum"(CCBK) van Brepols Publishers in Turnhout bedanken, dat het merendeel van de in dit proefschrift gebruikte digitale tekstedities ter beschikking stelde. Paul De Jongh, Bart Janssens, Jeroen Lauwers, en Luc Jocqué wil ik in het bijzonder bedanken om hun toewijding aan dit project. De leden van mijn doctoraatsbegeleidingscommissie, Francesco Stella en Maciej Eder, ben ik dankbaar om hun raadgeving doorheen mijn traject.

Aan CLiPS Computational Linguistics Group en Walter Daelemans: dank om de gastvrijheid waarmee ik werd ontvangen in de Lange Winkelstraat te Antwerpen begin 2017, een belangrijke periode voor de ontwikkeling van mijn methodologische behendigheden. Ik kijk ook met veel voldoening terug op mijn samenwerkingen met Latinisten Dinah, Klazina, Maxim, Stijn, Thomas en Tim in de organisatie van onderzoeksgroep RELICS en het tijdschrift JOLCEL. Dinah wil ik in het bijzonder bedanken om een klankbord te zijn geweest in mijn onderzoek naar Hildegard van Bingen, waaruit overigens een mooie samenwerking en een publicatie in *Parergon* is voortgekomen. Ook in deze thesis las Dinah het hoofdstuk van Hildegard na. Zij staat zo in het rijtje van andere *correctores* Enrique, Esther, Jeroen, Leen, Lisa, Marie-Gabrielle, Mathijs en Stefan die ik allen erkentelijk ben om hun grondige naleeswerk en inbreng. Caroline wil ik bedanken om de bijdragen die ze leverde als onderzoeksstagiar aan hoofdstuk 5. Tenslotte wil ik mijn dankbaarheid uiten aan de collega's van de vakgroep Geschiedenis in het UFO. Ik mag mezelf gelukkig prijzen dat ik vier jaar lang deel mocht uitmaken van dit bonte gezelschap. Een eervolle vermelding gaat uit naar mijn naaste collega's en vrienden van het Bureau D'hondt, Amber, Jirki, Johan, Leen, Lisa, Mathijs, Pieter, Tom en Ward. Ik kan niet genoeg benadrukken hoeveel de goede sfeer op het bureau me bemoedigd heeft in het schrijven van dit werk.

In mijn persoonlijke leven wil ik mijn lieve familie bedanken, mijn ouders Johan en Marie-Anne, mijn zus Nele en schoonbroer Bart. Mijn ouders hebben steeds hard gewerkt zodat ze hun kinderen de luxe van de keuze konden aanbieden. Hun onvoorwaardelijke steun is uitgemond in dit boek. Simon en Yann wil ik bedanken om de nachten in Gent; mijn bandmates Kim, Lorenzo en Maarten om de muzikale verstrooiing bij Coyote Melon.

Dat brengt me tenslotte bij mijn lief Esther. Niemands geduld werd het afgelopen jaar zo beproefd als het hare, en niemand deed zo veel opofferingen als zij. Het moet moeilijk zijn geweest voor haar, maar nu kan ze gerust zijn. Ondanks ons beider passie

voor het verleden, doet Esther me nu vooral dromen over de toekomst.

# Contents

# List of Figures

# List of Tables

# Introduction

## Research Aim

'Collaborative authorship,' as it features in this thesis's title, might have constituted a somewhat pleonastic compound to a medieval audience. After all, in the High Middle Ages, all authorship was intrinsically a collective endeavour. Authors were accustomed to working with secretaries, *librarii*, *notarii*, *amanuenses*, *scriptores*, *clerici*, calligraphers, ... and a variety of other specific roles and functions in composing, recording and archiving their works. Twelfth-century Latin writers such as Anselm of Canterbury (1033/4–1109), Guibert of Nogent (c. 1055–1124), William of Saint-Thierry (c. 1080–1148), Suger of Saint-Denis (1080/1–1151), Bernard of Clairvaux (1090–1153), Peter the Venerable (c. 1092–1156), Hildegard of Bingen (1098–1179), Elisabeth of Schönau (1129–1164/5), and many others, were accustomed to dictating their texts to a secretary, who would consequently entrust the text to parchment. Often this dimension goes lost when attributing and/or assigning a text to a medieval author by name, as often occurs in editions. We temporarily suspend our knowledge that the medieval author is in reality not reducible to a single name, but that (s)he is 'distributed' over an indefinite number of anonymous or withheld contributors who are left unnamed. In reality the medieval author is scattered and diffuse. Rooted in a tradition of the past, dependent upon the material conditions of the present, and subjected to alterations of the future, medieval authorship can not be narrowed down to one individual from the past, speaking to us in isolation and close confidence.

The collaborative aspect inherent to medieval authorship, and the implicit diversity of roles this 'collaboration' could denote, has for a long time posed historians and literary scholars for a difficult problem. Whereas it has been rightly acknowledged that such assistants were not merely instrumental in the literary process, but were active participants leaving a considerable mark on the image and style of the author and his or her text, this acknowledgement has nevertheless been accompanied by difficulties in defining the exact extent and sphere of influence of such secretarial mediation.

Five case studies of twelfth-century writing partnerships in Latin are investigated in this thesis, in which sensitive questions surrounding authority, synergy and gender take a prominent place. These case studies involve Bernard of Clairvaux and his secretary Nicholas of Montiéramey († 1176/8), Elisabeth of Schönau and her brother

Ekbert († 1184), Hildegard of Bingen and her biographers, Suger of Saint-Denis and his entourage, and the renowned twelfth-century lovers and correspondents Heloise of Argenteuil († 1164) and Peter Abelard (1079–1142). In exploring their collaboratively composed writings, we depart from questions on which aspects connect and/or distinguish them. Plenty of factors can be shown to vary from one case study of collaboration to the next: the material conditions, the practised genre, the authority of the figures involved, their education, sex, age, charisma, reputation, influence, etc.

In order to tackle the problem of twelfth-century collaborative authorship and its variety of expressions, this thesis departs from the premise that the identification of individual contributors to a medieval text —and the exact areas of their contributions— would constitute a particularly valuable insight. As of late, technical advancements in computer sciences have developed to the point of being able to facilitate such a need. Computational stylistics, the proposed methodology applied in this thesis, allows to make a statistically founded assessment of disputed texts' authorship(s) through the automatic detection of textual writing patterns.

The results brought forward in this thesis can be situated within two disciplines, abridged by the computational analysis of texts: medieval cultural history and literary studies. On a first level, it is argued that computational stylistics may provide a vital stepping stone in gaining more knowledge on the roles distributed amongst author(s), in gaining a better acquaintance with the historical context in which the author(s) operated, and of understanding various medieval models of literary composition.

Simultaneously, these computational experiments both contribute to and challenge the predominant historical and literary paradigms by which medieval authorship are studied, and allow to formulate new cultural-historical questions. Thereby we extend the usual application of stylometric methods beyond mere authorship attribution, and assess the practical and theoretical usefulness of computational stylistics as a new model for the analysis of medieval Latin authorship. This thesis's most important hypothesis is that computational stylistics provides a more accurate and objective means for capturing the 'distributionality' of twelfth-century authorship.[1] This term by which to describe authorship, inspired by the 'normal distribution' in statistics, embraces a phenomenon's core of individuality while acknowledging its mutability, captures its government both by laws and by deviations, and abides by typicalities in terms of trends without arguing for transfixion in categories.

[1] Also amongst historians and literary scholars I would not be the first to propose the usefulness of the term 'distributionality' in characterizing medieval authorship. See Bernadette A Masters, "The Distribution, Destruction and Dislocation of Authority in Medieval Literature and Its Modern Derivatives," *Romanic Review* 82, no. 3 (1991): 270–85.

# Historical Context

## The Twelfth Century

The problematization of 'collaborative' authorship needs to be seen in connection with the historical context in which this thesis's case studies are situated, where the matter of understanding and conceptualizing medieval authorship is a topic of ongoing debate for historians and literary scholars. This thesis is concerned with a particular time in the Middle Ages, and a particular type of literary culture within demarcated geographical boundaries. When I speak of medieval authorship, I will in effect be speaking exclusively of monastic and/or intellectual authors writing Latin literature in the geographical region enclosing parts of France and the Holy Roman Empire in the late eleventh- and twelfth centuries. The (long) twelfth century, famously termed a period of 'renaissance' by Charles H. Haskins in 1927,[2] is generally demarcated by a time span somewhere between 1054 (after the Great schism) and 1215 (the fourth Lateran council). In scholarly discussion, the twelfth century is often approached as a container of a vast number of 'changes' of a political, social and cultural nature, captured under such terms as 'renewal,' 'revival,' 'reform,' etc.

Indeed, the breadth of the century's developments is impressive. So, for instance, the education system branches out from the monastic schools towards the cathedral schools and independent institutions that would ultimately become full-fledged universities.[3] This shift in how knowledge was taught and the arrival of a new audience to benefit from it, is most evidently demonstrable in Paris. Typically one will find the spirit of this transformation best embodied in theologian Peter Abelard. Simultaneously, the church was coping with gradual yet profound clerical transformations in the wake of Gregorian reform.[4] New monastic ideals proliferated, advocating new ascetic ideals and new forms of cenobitism and seclusion. One could think of Bruno of Cologne's (c. 1032–1101) Carthusian order, Norbert of Xanten's (1080–1134) Premonstratensians, but best known are the Cistercians who found their most successful promoter in abbot Bernard of Clairvaux.[5] All kinds of other cultural-historical diversifications come in an unprecedented scope. We find the origins of a new visionary literary tradition in the Rhineland with Hildegard of Bingen, wielding a type of authorship and authorial influence that would come to have specific gendered associations

[2] Charles Homer Haskins, *The Renaissance of the Twelfth Century*, 8th ed. (Cambridge, MA: Harvard University Press, 1971 (1927)).

[3] A seminal work on the rise of eleventh- and twelfth-century cathedral schools is C. Stephen Jaeger, *The Envy of Angels. Cathedral Schools and Social Ideals in Medieval Europe 950–1200*, Middle Ages Series (Philadelphia: University of Philadelphia Press, 1994).

[4] The 'Gregorian' reform, named after its main promoter Pope Gregory VII (c. 1015–1085). The appearance of Augustin Fliche's classic, in which the term 'Gregorian reform' was coined, advanced Gregory VII and his circle as the central personality of the movement. See Augustin Fliche, *La réforme grégorienne* (Louvain: Spicilegium sacrum Lovaniense, 1924–37).

[5] Giles Constable, *The Reformation of the Twelfth Century* (Cambridge, UK: Cambridge University Press, 1996).

in the centuries to come. With Suger of Saint-Denis's (1081–1151) rebuilding of the basilica of Saint-Denis we observe the origins of Gothic aesthetics. Elsewhere, writers and poets rediscover Platonic texts —especially Cicero's translation of *Timaeus*—, which causes the instigation of distinctive schools such as the school of Chartres, bringing forth influential writers Bernard Silvestris (1085–1160/78), William of Conches (1090–1154) and Alan of Lille (c. 1120–1202).[6] Especially in the second half of the century we see the first signs of the vernacularization of literature —especially in courtly romances— falling into rivalry with the dominant Latin tradition, most notably with such figures as Marie de France (1160–1215) and Chrétien de Troyes (c. 1140-1190). In philosophical, theological and scientific learning we find exchanges between European and Arabic thought, with interpreters such as Hermann of Carinthia (c. 1100–1160) or Adelard of Bath (c. 1080–1152). In Bologna, Roman law is rediscovered and applied in Gratian's *Decretum Gratiani*, etc.

This all merely serves to indicate that the twelfth century is an age 'on the move,' and moving subjects are harder to grasp. In this context of many cultural developments, the twelfth century has often been deemed transitional and even 'experimental' in many facets. This same elusiveness applies for the study of twelfth-century authorship. In what is to follow, we will discuss how an interesting dilemma arises in how twelfth-century authors of Latin required the 'participation' of others (which can safely be broadly interpreted), whilst increasingly cultivating an expression of 'self.'

## The New Philology

The Middle Ages is a manuscript culture, in which texts were unproblematically filtered or renegotiated through oral transmission, the mediation of assistants and scribes, and incessant borrowing and/or adaptation (no reference required) from authoritative sources. These physical conditions strongly impacted the ways in which the author practised his or her métier, and automatically caused appropriation and authority over a text to be a much more flexible given than in modern times. This idea was most famously advocated by the New Philology movement, gaining momentum in the 1990s after the publication of the influential 1990 *Speculum* 65.1 issue. In the issue's programmatic introduction, Stephen G. Nichols would advocate the "desire to return to the medieval origins of philology,"[7] which essentially meant that analyzing and un-

[6] Winthrop Wetherbee, *Platonism and Poetry in the Twelfth Century. The Literary Influence of the School of Chartres* (Princeton: Princeton University Press, 1972).

[7] Stephen G. Nichols, "Introduction: Philology in a Manuscript Culture," *Speculum* 65, no. 1 (1990): 1; the heritage of post-structuralists such as Roland Barthes (1915–80) and Michel Foucault (1926–84) has often been noted. See Roland Barthes, "The Death of the Author," *Aspen* 5–6 (1967); and Michel Foucault, "Qu'est-ce qu'un auteur?," *Bulletin de la société française de philosophie* 69, no. 3 (1969); also see Virginie Greene, "What Happened to Medievalists After the Death of the Author?," in *The Medieval Author in Medieval French Literature*, ed. Virginie Greene, Studies in Arthurian and Courtly Cultures (New York: Palgrave, 2006), 205–27; and Jeroen Deploige, "Anonymat et paternité littéraire dans l'hagiographie des Pays-Bas Méridionaux (*ca.* 920 - *ca.* 1320). Autour du discours sur l'«original» et la «copie» hagiographique au Moyen Âge," in *« Scribere sanctorum gesta »: Recueil d'études d'hagiographie médiévale offert à Guy*

derstanding all the various physical appearances of written documents is an essential hermeneutic criterium for understanding both medieval text and author. Due to this emphasis on materiality, the New Philology's alternative name, likewise launched by Nichols, was the Material Philology.[8] The movement drew attention to how a text's meaning is heavily determined by the social embedding in which it was materially composed and in which it circulated. The medievalist should embrace the notion that every variant of the text is an artifice, worthy of study on its own, and a result of a diachronic, organic and open-ended composition process in which multiple actors such as secretaries, scribes and copyists were involved. This fluctuation of the medieval text —or 'mouvance,' in Paul Zumthor's terms—[9] is to be praised instead of hierarchized.

The ideas of the New Philology gradually debunked the Lachmannian framework in which a canonical author or canonical 'Urtext' are deemed recoverable or desirable.[10] This lack of stability of author and text illustrated that interpreting medieval literary production as the outpouring of a single individual's inspiration is untenable. The medieval author is no romantically inspired genius, who silently entrusts his or her thoughts to parchment in a solitary setting. Medieval text is always 'collaborative' in the broadest sense of the word.

The New Philology's main point of focus was reception and appropriation when the text entered the phase of transmission, which allowed it to appear in various forms or was granted new meaning through its accompaniment by additional texts. Yet, one may well emphasize how textual transference and transformation was already present in the early phases of conception, when assistants and secretaries first wrote down what was dictated. As will be rehearsed time and again throughout this thesis, medieval text originated in an oral-aural setting, where not only the author and his or her secretaries, but also a tradition and a sacred language dictated the words to be inked in the manuscript. Once this copy was finished, there was always the future —with its potential amendations— to be reckoned with. From its first draft to its 'final' state of redaction, not ever can we safely presume that we are reading the literary product of a single individual. In an age before the invention of the printing press, a text could remain 'open-ended' instead of having a final state of publication, a context in which it has often been argued that our modern-day conception of plagiarism or text-theft was non-existent.[11] In his seminal book *Éloge de la variante*, which lay at the basis

*Philippart*, ed. Étienne Renard et al. (Turnhout: Brepols, 2005), see especially 84.

[8] Stephen G. Nichols, "Why Material Philology? Some Thoughts," *Zeitschrift für deutsche Philologie* 116 (1997): 10–30.

[9] The concept of 'mouvance' was formulated in the second chapter of Zumthor's study of medieval French poetry, see Paul Zumthor, *Essai de poétique médiévale* (Paris: Éditions du Seuil, 1972), 65.

[10] Karl Lachmann (1793–1851) and his edition of Lucretius' *De rerum natura* laid the foundations for nineteenth-century German philology. Lachmann's idea was that an agreement between errors in manuscripts implied their common origin. In other words, he devised that a hierarchy of texts can be established according to origin, leading up to an archetype or urtext that is preferable to the others. See Titus Lucretius Carus, *De rerum natura libri sex*, ed. Karl Lachmann (Berlin: George Reimer, 1850).

[11] The inadequacy of terms such as forgery and plagiarism is discussed more extensively on p. 48 in this thesis.

of the New Philology movement, these observations brought Bernard Cerquiglini to the conviction that "the author is not a medieval concept."[12]

## The Rise of Individualism?

And yet, simultaneously, for scholars such as Alastair Minnis, the twelfth-century author is the individual author in potency. The twelfth century indeed testifies of a type of literature in which we observe increased self-reflection and introspection. This in its own turn goes accompanied by a text type that is oriented upon the writing subject, and raises the impression of autobiographical accounts. The best examples of this early process are ego-narratives such as Peter Abelard's *Historia Calamitatum*, or those of Guibert of Nogent in his *Monodiae*, but also in the German sphere one can point to the fragmentary autobiographical passages in Hildegard of Bingen's visionary texts, or Rupert of Deutz's personal presence in the reading and interpreting of the Bible.[13] Thereby, it already forms witness of —or is the germination process of— what arrives in the thirteenth century with such authors as the Franciscan theologian Bonaventura (1221–74).

In Minnis's fundamental work on medieval authorship this is described as the "emergence of an inspired but fallible author who was allowed his individual authority and his limitations, his style and his sins."[14] He takes as point of departure the study of the prologues of commentaries and exegetical works especially from the later scholastic period, and describes how the schools defined for themselves a framework of literary theory from the newly translated Aristotelian logic, through which they could approach the biblical texts and patristic *auctores* more literally, and therefore more literarily, as "a new type of exegesis emerged, in which the focus had shifted from the divine *auctor* to the human *auctor* of Scripture."[15]

We find other attestations in scholarly discussion proclaiming the twelfth century "discovering the individual."[16] Accompanied with this discovery of the individual

[12] "L'auteur n'est pas une idée médiévale," see Bernard Cerquiglini, *Éloge de la variante: Histoire critique de la philologie* (Paris: Seuil, 1989), 25; The book was first published in French in 1989, and was based on an earlier article Bernard Cerquiglini, "Éloge de la variante," *Langages* 17, no. 69 (1983): 25–35; the (often cited) English translation followed ten years after as Bernard Cerquiglini, *In Praise of the Variant: A Critical History of Philology*, 2nd ed., trans. Betsy Wing, Parallax: Re-Visions of Culture and Society (Baltimore: The John Hopkins University Press, 1999 (1989)), 8.

[13] "He constructs an identity in and through the explication of Scripture as personal experience," see Morgan Powell, "*Vox ex Negativo*. Hildegard of Bingen, Rupert of Deutz and Authorial Identity in the Twelfth Century," in *Unverwechselbarkeit. Persönliche Identität und Identifikation in der vormodernen Gesellschaft*, ed. Peter von Moos, Norm und Struktur 23 (Cologne: Böhlau Verlag, 2004), 282.

[14] Alastair J. Minnis, *Medieval Theory of Authorship: Scholastic Literary Attitudes in the Later Middle Ages*, 2nd ed., The Middle Ages Series (Philadelphia: University of Pennsylvania Press, 1988 (1984)), xiii.

[15] Ibid., 5.

[16] See Colin Morris, *The Discovery of the Individual, 1050-1200* (Toronto: University of Toronto Press, 1987 (1972)); Caroline Walker Bynum, "Did the Twelfth Century Discover the Individual?," *Journal of Ecclesiastical History* 31, no. 1 (1980): 1–17; Colin Morris, "Individualism in Twelfth-Century Religion. Some Further Reflections," *The Journal of Ecclesiastical History* 31, no. 2 (1980): 195–206; Aron Iakovlevič Gurevič and Katharine Judelson, *The Origins of European Individualism*, trans. Katharine Judelson, The Making of Europe (Oxford: Blackwell, 1995).

comes an increased literacy and importance of written record,[17] and a taste for leisurely reading,[18] all of which flourishing also outside of the monastery walls in administrative centres, secular institutions and the courts. Literacy was expanding from the closed and the contemplative to the open and the entrepreneurial. Some of the most charismatic and remarkable authors of the Middle Ages, such as Peter Abelard and Bernard of Clairvaux, come forward. In spite of the New Philology's claims of a material culture which allowed no leeway to concepts as individual expression, copyright and text ownership as it is known in a modern print culture, some sense of individuality and expression of the self appears to have been on the rise nevertheless.

Whether or not one can truly speak of the rise of individualism in the twelfth century has been a topic of long discussion amongst medievalists. As noted by Bynum,[19] one should be careful to equate twelfth-century forms of individuality all too lightly with its modern, contemporary understanding. Although one may encounter terms such as 'renaissance' or 'humanism' being tossed around in reading through this scholarly debate, quite a few medievalists have shown unease with such words' incompatibility and somewhat anachronistic implications. Despite the rediscovery of the classics in the school curriculum and a sharpened sense of what one might call 'individuality' or 'self,' the twelfth century does not as yet bring us any Shakespeares.[20]

## A Theoretical Dilemma

As the foregoing paragraphs suggest, the twelfth century with its well-known range of cultural shifts does not always fit well into theoretical straitjackets. In particular, not only do we have a medieval author as (s)he is presented by the New Philology, divided and subjected to the material conditions, we also have the gradual arrival of an 'individualizing' medieval author as presented in the previous paragraph.

As will be a main point of exploration in chapter 1, this tension field has an impact on our understanding of collaborative writing. On the one hand, authors relied upon professional scribes and secretaries which allowed them to spare time and effort. On the other, authorship is increasingly related to the expression of a "new intimacy linking author, text, and reader."[21] In short: authors sought the participation of others,

[17] Brian Stock, *The Implications of Literacy: Written Language and Models of Interpretation in the Eleventh and Twelfth Centuries* (Princeton: Princeton University Press, 1983); Michael T. Clanchy, *From Memory to Written Record, England 1066–1307* (Oxford, UK: Blackwell, 1993 (1979)).

[18] Wim Verbaal, "How the West was Won by Fiction: The Appearance of Fictional Narrative and Leisurely Reading in Western Literature (11th and 12th century)," in *True Lies Worldwide: Fictionality in Global Contexts*, ed. Anders Cullhed and Lena Rydholm (Berlin: De Gruyter, 2014), 189–202.

[19] See Bynum, "Did the Twelfth Century Discover the Individual?"

[20] William A. Nitze, "The So-Called Twelfth-Century Renaissance," *Speculum* 23, no. 3 (1948): 464–71; Bynum, "Did the Twelfth Century Discover the Individual?"; Ineke van 't Spijker, *Fictions of the Inner Life. Religious Literature and Formation of the Self in Eleventh and Twelfth Centuries*, Disputatio 4 (Turnhout: Brepols, 2004), see especially 3.

[21] Paul Saenger, *Space Between Words: The Origins of Silent Reading*, Figurae: Reading Medieval Culture (Stanford, CA: Stanford University Press, 1997), 243.

whilst increasingly cultivating an expression of 'self.' More writers are externally involved, and yet internally the text presupposes fewer voices than before. Accompanied with this disjunction between intellectual author and physical scriptor, we observe an increasing concern over the loss of authorial control. We find attestations where value is attached to titled authority, where there is an outspoken preference for unviolated text, or where personal literary style is cultivated.

This presents a tension field in which it is not always easy to orientate. Nichols, when discussing the tendency in medieval culture to be both "mutable" and yet "resist change in and for itself," addressed this dilemma as the "paradox of medieval culture: its delicate and seemingly contradictory balance between stability, on the one hand, and transformation, on the other."[22] Even though Nichols was mainly speaking of texts in the vernacular rather than authors writing Latin, where one may say matters are —to a modern point of view— more or less 'under control,' it is a remark that captures quite accurately the theoretical dilemma we wish to address. On the one hand we see the composition of text as shared, composed in mimesis with authoritative forebears and with assistance of collaborators: a dispersed author. On the other we see the rise of an individual author who is increasingly aware of his or her subjectivity and perceives of the text as an expression of the self: a centralized author. In the individual case studies of this thesis, already briefly lined up above and more elaborately introduced below, this perceptible yet undefined distribution of roles at work under the surface raises a whole range of sensitive questions on the nature of the text, its composition history and its authorship.

## Computational Stylistics

Computational stylistics or stylometry is a subfield of the digital humanities (DH), based on techniques from computational linguistics and natural language processing (NLP), in which computational methods are applied to segregate writing styles. A more thoughtful definition of computational stylistics is proposed in the beginning of chapter 2, but for now we can limit ourselves to defining the method by what it does best, and by the purpose for which it will predominantly be put to use: authorship attribution.

The basic yet quite revolutionary idea of computational stylistics is that writers have individual writing habits which can be statistically traced within the text.[23] What is indeed quite fascinating about the field is that it has been able to prove that writers consistently betray linguistic characteristics in their language use which are highly per-

[22] Stephen G. Nichols, *From Parchment to Cyberspace. Medieval Literature in the Digital Age*., Medieval Interventions. New Light on Traditional Thinking 2 (New York: Peter Lang, 2016), 56.

[23] Efstathios Stamatatos, "A Survey of Modern Authorship Attribution Methods," *Journal of the American Society for Information Science and Technology* 60, no. 3 (2009): 538–556.

sonal, a sort of stylistic DNA (or 'stylome').[24] Computational stylistics' success lies in its ability to automatically detect stylistic features that are subconsciously processed, have no semantic content and yet occur in high frequency, so that they escape the attention of even the most skilled imitators. Popularly applied in this regard are function words, insignificant marker words with a syntactic and grammatical function such as adverbs, conjunctions, prepositions, pronouns, articles, auxiliary verbs, ... As we will come to show below, these are excellent predictors of authorship in synthetic or highly inflected languages such as Latin as well.

On a first level, text attribution by means of computational stylistics is a way of providing a text its context. Acquainting ourselves with a text's truthful author allows us to get a clearer grasp on its intentions and enhance our understanding of it. Indeed, most questions in this thesis will recall a well-known, centuries-old historical and philological fascination for classifying anonymous documents, and segregating what is 'authentic' from what is 'pseudepigraphic' or 'falsely ascribed.' This has always been a matter of concern for medieval scholars, to detect false ascriptions, forgeries, corruptions or interpolations, those encumbrances which the diplomatist and Maurist Jean Mabillon (1632–1707) called "evils hiding in plain sight,"[25] and which the New Philologists came to embrace as manifestations of variance. Considering that specialized disciplines for detecting these such as textual criticism, palaeography, diplomatics and codicology largely rely on physical, text-external characteristics rather than text-internal ones, an additional (auxiliary?) discipline for assessing a text's authorship where such physical evidence is lacking seems most welcome. In other words, computational stylistics provides a novel approach to *Echtheitskritik*. But this is only the first step.

As indicated by the theoretical dilemma elaborately introduced earlier, maintaining a model in imitation of nineteenth-century forms of 'Echtheitskritik,' and classifying texts according to either 'A' or 'B' is undesirable, and I hope to have convinced the reader by the end of this book that computational stylistics can indeed be applied to establish more than merely reinvigorate existent categories. In many cases it succeeds quite well at questioning them, establishing new ones, or simply presenting medieval authorship in the 'variance' that typifies it. In this light one may emphasize how Nichols and Cerquiglini, both influential figures for the New Philological idea as introduced above, were drawn to the computer's abilities to explore medieval 'mutability.'[26] Cerquiglini concluded his influential monograph by stating that the

[24] Hans van Halteren et al., "New Machine Learning Methods Demonstrate the Existence of a Human Stylome," *Journal of Quantitative Linguistics* 12, no. 1 (2005): 65–77; Mike Kestemont, "Function Words in Authorship Attribution: From Black Magic to Theory?," in *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL)* (Gothenburg, Sweden, April 2014), 59–66.

[25] "Sed inquirendum, quam late pateat hoc malum, et si qua tandem arte ipsi occurri possit," see Jean Mabillon, *De re diplomatica libri VI*, 2nd ed. (Paris, 1709 (1681)).

[26] Although they did not necessarily have computational stylistics in mind, but rather the computer's ability to digitize, store a large amount of data, and access documents simultaneously at all times, these intuitions may well be extended into

"computer, a valuable aid and one worth considering, provides the obvious solution" for capturing the full variety of medieval literary production. Indeed, studies by Kestemont,[27] Moens, Deploige,[28] van Dalen-Oskam,[29] van Zundert,[30] and myself[31] have given indication of how computer-assisted analyses of style may delineate the individual contributions made by scribes, secretaries, and authors within medieval text.[32] The aim is not to compartmentalize, but merely to explore the composite character of medieval literature, and map out the respective spheres of influence.

It is easily forgotten that statistics are actually better equipped than most fields with an evaluative terminology to give expression to doubt or to a degree of confidence through ranking, percentages, gradations, geometrical spectra and schematic visualizations. Indeed, instead of making binary choices between 'A' or 'B,' we enter the realm of probabilities and distributions.

One may note in this regard that this thesis does not unidirectionally challenge the Middle Ages by using modern techniques. The opposite also holds true: the Middle Ages pose modern techniques before methodological challenges in their own turn.[33] Although one may enumerate quite a few such challenges,[34] medieval authorship's collaborative aspect may be counted amongst the most substantial. Previous stud-

that field. Computers indeed have the advantage of comparative insight and in-depth analysis of complex and multivariate patterns. See Cerquiglini, *In Praise of the Variant*, 79; and Nichols, *From Parchment to Cyberspace*.

[27] See especially chap. 3 on Middle Dutch copyists in Mike Kestemont, *Het gewicht van de auteur. Stylometrische auteursherkenning in Middelnederlandse literatuur*, Studies op het gebied van de oudere Nederlandse letterkunde 5 (Gent: Koninklijke academie voor Nederlandse taal- en letterkunde, 2013); Mike Kestemont, "Stylometry for Medieval Authorship Studies: An Application to Rhyme Words," *Digital Philology: A Journal of Medieval Cultures* 1, no. 1 (2012): 42–72.

[28] Mike Kestemont, Sara Moens, and Jeroen Deploige, "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux," *Digital Scholarship in the Humanities* 30, no. 2 (2013): 199–224.

[29] Karina van Dalen-Oskam, "The Secret Life of Scribes. Exploring Fifteen Manuscripts of Jacob van Maerlant's *Scolastica* (1271)," *Literary and Linguistic Computing* 27, no. 4 (2012): 355–72; Karina van Dalen-Oskam, "Authors, Scribes, and Scholars. Detecting Scribal Variation and Editorial Intervention via Authorship Attribution Methods," in *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, ed. Tara Andrew and Caroline Macé (Turnhout: Brepols, 2014), 141–57; Mike Kestemont and Karina van Dalen-Oskam, "Predicting the Past: Memory Based Copyist and Author Discrimination in Medieval Epics," in *Proceedings of the Twenty-First Benelux Conference on Artificial Intelligence*. Ed. Toon Calders, Karl Tuyls, and Mykola Pechenizkiy (Eindhoven: BNVKI-AIABN, 2009), 121–8.

[30] Karina van Dalen-Oskam and Joris van Zundert, "Delta for Middle Dutch - Author and Copyist Distinction in Walewein," *Literary and Linguistic Computing* 22, no. 3 (2007): 345–62.

[31] Jeroen De Gussem, "Bernard of Clairvaux and Nicholas of Montiéramey: Tracing the Secretarial Trail with Computational Stylistics," *Speculum* 92, no. S1 (2017): S190–S225.

[32] By which I leave out of regard computer-assisted stemmatology or the making of digital editions of medieval texts, which may in some cases be closely related. See Peter Robinson, "The Digital Revolution in Scholarly Editing," in *Ars Edendi Lecture Series, vol. IV*, ed. Barbara Crostini, Gunilla Iversen, and Brian Møller Jensen, Studia Latina Stockholmiensia (Stockholm: Stockholm University Press, 2016), 181–207; Barbara Bordalejo, "The Genealogy of Texts: Manuscript Traditions and Textual Traditions," *Digital Scholarship in the Humanities* 31, no. 3 (2016): 563–577; Tara L. Andrews and Caroline Macé, "Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation through Graph Analysis of Texts and Stemmata," *Literary and Linguistic Computing* 28, no. 4 (2013):

[33] Medievalists have been considered "early adopters"of the computer. See John Unsworth, "Medievalists as Early Adopters of Information Technology," *Digital Medievalist* 7 (2012).

[34] The challenges posed by computational stylistics to medieval literature and vice versa will be a topic of discussion throughout chapters 2 and 3.

ies applying computational stylistics to writings of mixed authorship have indicated specific problems, some of which yet unresolved.[35] It stands beyond doubt that determining if one and the same text was written by a single or multiple authors is a more difficult task than selecting either class 'A' or 'B.' For instance, all of these studies have reported on finding that collaborative forms of authorship may result in a style that is unlike each of the potential candidates to have contributed to a text.[36] The current state of the art is still actively looking to improve its methods in this domain. A glance at the most popular tasks within the annual PAN competition indicate that detecting style change within a text is proposed as one of the harder problems to solve.[37]

# Design

## General Structure

The first part of this thesis, comprising chapters 1 to 3, is of a theoretical-methodological nature. It mainly serves to introduce in full the historical problem of twelfth-century collaborative authorship at hand (chapter 1), and to give a methodological exploration of computational stylistics (chapters 2 and 3). Computational stylistics is first carefully defined and embedded in a long-standing philological tradition (chapter 2), before its practical usefulness for attributing twelfth-century Latin texts is displayed by hands-on examples (chapter 3).

The second part of this thesis, comprising five chapters (4–8), contain the case studies. The case studies do not follow a chronological order. Instead, they are arranged in such a manner that the thematic emphasis gradually shifts from collaborative authorship in its material, physical form, to collaborative authorship as literary device, in which authorial role-playing is deliberately modelled. The first predominantly seeks to map out the dynamics of the author-collaborator relationship and multiple authors' stylistic contributions to a text, whereas notions such as impersonation, forgery and fictionality gradually play a larger part in the final pages. One could argue in this

[35] Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century"; Karina van Dalen-Oskam, "Epistolary Voices. The Case of Elisabeth Wolff and Agatha Deken," *Literary and Linguistic Computing* 29, no. 3 (2014): 443–51; Alexander A. G. Gladwin, Matthew J. Lavin, and Daniel M. Look, "Stylometry and Collaborative Authorship: Eddy, Lovecraft, and 'The Loved Dead'," *Digital Scholarship in the Humanities* 29, no. 3 (2014): 443–51.

[36] Pennebaker's 'synergy hypothesis' may come to mind, as proposed in Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 214; Pennebaker originally described it as a situation "when two people work closely together, they create a product unlike either of them would on their own. Their language style will be distinctive in a way such that most people would not recognize who the author was," see James W. Pennebaker, *The Secret Life of Pronouns. What Our Words Say About Us* (New York: Bloomsbury Press, 2011), 553.

[37] This is a series of scientific events and online shared tasks on text forensics and stylometry, co-located with the annual CLEF conference (International Conference of the Cross-Language Evaluation Forum for European Languages). The idea is that difficult problems for computational stylistics are solved in a community sharing code and data to advance the field. Software is submitted, and the best-performing model wins the competition. For an overview of last year's topics of interest, see Efstathios Stamatatos et al., "Overview of PAN 2018," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, ed. Patrice Bellot et al. (Cham: Springer International Publishing, 2018), 267–85.

regard that the thesis ascends from the reality of the scriptorium to the reality of the text, which gradually clears the stage to problems characteristic for literary rather than historical studies. This kind of progression allows a piecemeal acquaintance with the thesis's methodology and research problem.

Throughout this thesis I will from time to time refer to the appendix found at the back of this thesis (from p. 304 onward). The appendix mostly contains technical details or additional data that may be of interest in interpreting the results given throughout the chapters, but does not contain vital information for understanding the arguments and conclusions. Each chapter in this thesis has an individual section in the appendix, coded as 'A.{x},' where {x} is substituted by the chapter number.

## Case Studies

The selected case studies in this thesis (chapters 4–8) are ordered according to the rationale of the thematic sliding scale as presented earlier, which runs from the reality of the scriptorium to the reality of the text. Although broader connections will be drawn between the different cases discussed, each of them places in the forefront one specific twelfth-century figure. What binds all of these authors is that they were (broadly speaking) contemporaries, operating within a religious monastic context (and not seldom school context) in France and the Holy Roman Empire. They all predominantly wrote their texts in Latin, which at the time was the main language of the literate, at the court, in the Church and at the schools. Naturally, these authors are also bound by the quality of having composed their works in a collaborative setting. It deserves emphasis that some of them would have known each other, and some of them corresponded or even encountered one another in real life, on friendly but also on hostile terms. All of these aspects will become clearer as we dig deeper into the twelfth century and discover each of these figures in some more detail. Here I will give each of them a concise introduction.

### Bernard of Clairvaux

The first author which we will come to discuss in chapter 4 is the French abbot Bernard of Clairvaux (c. 1090-1153). Bernard was the figurehead (not the founder) of the Cistercian movement, and an indispensable figure for understanding twelfth-century monastic reform. Having founded Clairvaux Abbey in 1115, Bernard gradually took central stage in Church politics, and entertained a wide network of powerful correspondents. One particularly well-known relationship is that with abbot Peter the Venerable of Cluny. Peter the Venerable, that other monastic giant of the twelfth century, represented the traditional type of Benedictine monasticism which Bernard criticized. In 1125, Bernard wrote his polemic *Apologia*, which became a monumental document denouncing the intemperance and excesses of Cluniac monastic life. He made a further name for himself by settling the papal schism in favour of Pope Inno-

cent II against Antipope Anacletus (between 1130 and 1138), his theological dispute with the Parisian school master Peter Abelard at the Council of Sens in 1141 (discussed below), and his preaching on the failed Second Crusade (1147–1149).

As a writer, Bernard boasts a rich oeuvre. He is known to have been a brilliant and charismatic preacher, and the literary style of his writings has often been highly praised. The eighty-six *Sermones super Cantica Canticorum* (*Sermons on the Song of Songs*) are considered an absolute masterpiece of the Middle Ages and a highlight of Latinity. Bernard was an engaged and occupied figure of statute and authority, a man of letters who strove to convince his audience by composing beautiful texts. Simultaneously, Bernard is known to have delegated his work to secretaries, some of which were exceptionally professional and well-educated in rhetoric and epistolography. Their main skill was to imitate Bernard and write drafts of letters and sermons in his person from his dictations. Such secretaries' mediation on several particular texts within Bernard's corpus, and more generally on his entire oeuvre, has fallen subject to much debate. At stake was Bernard's authority and reputation, which —as his many occupations increased— lessened his grasp on the outgoing correspondence, a matter which he grew increasingly sensitive to. In chapter 4, we will discuss the impact one such professional secretary in the Clairvaux scriptorium, namely Nicholas of Montiéramey, may have had on the final outlook of Bernard's oeuvre.

### Elisabeth of Schönau

With the visionary and Benedictine nun Elisabeth of Schönau (1129–1164/5) we move from France to the Holy Roman Empire and find ourselves in the double monastery of Schönau in the diocese of Trier.[38] The little we know of Elisabeth's background suggests that she was a descendant of a line of Saxon nobility residing in the area of Cologne and Bonn. At the age of twelve, she was oblated to the monastery of Schönau for unknown reasons.[39] When she reached the age of twenty-three in 1152, Elisabeth began to experience her first divinely inspired visions, and was encouraged by abbot Hildelin to persist. Aided for some time by the nuns in the convent, she was soon accompanied by her brother Ekbert († 1184) some three years later, who became her assistant in recording, translating and revising her visions. Ekbert was a former deacon at Saint Cassius in Bonn, and the reasons why he chose a monastic life by his sister's side in Schönau over a career in the Church are not clear. It appeared to have raised a few eyebrows with his contemporaries, as the latter (more prestigious) path certainly would have been open to him. Elisabeth was no Bernard of Clairvaux

[38] For the different types of male-female symbioses covered by the term 'double house,' see Kaspar Elm and Michel Parisse, *Doppelklöster und andere Formen der Symbiose männlicher und weiblicher Religiosen im Mittelalter*, Berliner historische Studien. Ordensstudien 8 (Berlin: Duncker und Humblot, 1992).

[39] Kurt Köster, "Elisabeth von Schönau. Leben, Persönlichkeit und visionäres Werk," in *Schönauer Elisabeth Jubiläum 1965. Festschrift anläßlich des achthundertjährigen Todestages der heiligen Elisabeth von Schönau* (Limburg: Pallotiner Druckerei, 1965), 20–1.

when it comes to fame and influence, and yet her visions not seldom seem to be concerned with topical concerns and activities in the Church institute at large. Both she and Ekbert maintained a network of predominantly local, high-placed members of the clergy. So Elisabeth's visions give unique insights into the Church's contemporary concerns over a heretical movement which came to be termed 'Catharism,' —a word which Ekbert himself may have been the instigator of—, and it appears that one of Elisabeth's revelations may have served to 'authenticate' suspect relics for the sake of winning the favour of the abbot of Deutz.

One cannot but understand Elisabeth's texts and the visionary tradition she is part of without referring to Hildegard of Bingen, her older and better known paragon to be discussed below. Elisabeth corresponded with Hildegard, and was patently influenced by her. Nevertheless, scholars have based themselves on the available manuscript evidence to maintain that Elisabeth may well have been more popular than Hildegard in her own time.

When it comes to the composition of their visions, one must note that there were notable differences between the two visionaries. Apparently Elisabeth occasionally spoke her visions in Latin, but also uttered them in her mother tongue, or even mingled both languages. That Hildegard dictated her visions in any other tongue than Latin has not been attested. This enhances the impression —and it may remain but an impression— that Elisabeth's Latin would have been extremely inadequate, perhaps more so than Hildegard, for which reason she heavily relied on her brother for most of the phrasing. Ekbert's intrusions in the texts are so frequent that it has often been assumed that the latter played a larger part in redacting and revising his sister's works than Hildegard's secretaries, and was much more at liberty to alter Elisabeth's wording. We will come to see in chapter 5 that what Elisabeth's text purports about the method in which she composed her texts and what actually took place may well be two different realities. We will also observe that the difficulties in interpreting Elisabeth's self-devaluations and Ekbert's self-inclusion are symptomatic for cross-sex collaborations, where such characterizations often fulfill a topical function and comply to gendered associations.

### Hildegard of Bingen

The renowned twelfth-century visionary Hildegard of Bingen (1098–1179), the 'Sibyl of the Rhine,' was born in the region of Mainz from a family of noble lineage, and started religious life as an oblate at the age of eight. What exactly occurred after this time is subject to debate,[40] but it is certain that some years later she took her vows and

[40] The primary sources allowing to build up a timeline of Hildegard's childhood years are contradictory. What exactly happened with Hildegard between her eight year and her enclosure in Disibodenberg (which could impossibly have taken place in 1106, since construction at Disibodenberg only took place in 1108) remains subject of debate. One theory holds that Hildegard and Jutta were taught in Sponheim by the widow Uda of Göllheim before they took vows in Disibodenberg. For the most recent state of affairs, see Franz J. Felten, "What Do We Know About the Life of Jutta and Hildegard at

was enclosed in the double house of Disibodenberg as a recluse with anchoress Jutta von Sponheim (1091–1136), her six years older teacher. The enclosure in Disibodenberg entailed the incarceration in a cell from which she could communicate with visitors during fixed hours and received life supplies through a slit in the wall. After Jutta's death in 1136, Hildegard was elected as the prioress of what had gradually become a small female community.

It was several years later, in 1141, that Hildegard confided to a monk named Volmar that she was visited by the Holy Spirit in visions, and possessed of a divine prophetic gift. With abbot Kuno's approval she redacted *Scivias* that same year, the first treatise of what was to become her well-known visionary trilogy, additionally consisting of *Liber vitae meritorum* and *Liber divinorum operum*. A life-turning event occurred in 1150 when Hildegard disconnected from Disibodenberg, to the annoyance of some of its members, and moved to the Rupertsberg in Bingen with a small band of female followers and her provost Volmar. It was here, somewhat remote from the Disibodenberg cloister, that Hildegard would work on the remaining of her rich oeuvre, and composed music, plays, medicinal works, scientific treatises, and letters. There she appears to have attracted the gaze of popes, bishops and Parisian teachers, all of whom were fascinated by the appearance of a compelling female figure by the Rhine, whose authority was based on an informal and charismatic relationship to God.

As with Elisabeth, Hildegard was flanked by male and female collaborators such as Volmar of Disibodenberg († 1173), Godfrey of Disibodenberg († 1175/6), Guibert of Gembloux (1124/25–1214) and Richardis of Stade in the composition of her works. She conventionally presented these collaborations to her readership as a necessity due to her limited grasp over the language. Nevertheless, Hildegard manifestly warned her assistants not to change the sense of her visions, and to merely revise formal aspects of the language such as grammar and spelling. It is unlikely that Hildegard truly was as 'indocta' as she tends to let on. As Gössmann stated, "not only in Hildegard's but also in other medieval women's writing, female formulas of humility and the theme of womanhood as weakness become an expression of strong self-confidence since humiliation is answered by God's calling and exaltation."[41] Her works testify of an encounter with patristic thought, be it presumably to a larger extent in an indirect fashion (liturgy, correspondence, anthologies, loose quotations, ...), and she proclaims to have been familiar with some of the classical authors, amongst whom she may have explicitly named Donatus and Lucan in a fragmentary autobiographical passage, which is nevertheless surrounded by spurious *sententiae*.[42]

Disibodenberg and Rupertsberg?," in *A Companion to Hildegard of Bingen*, ed. Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, Brill's Companions to the Christian Tradition 45 (Leiden: Brill, 2014), 15–38.

[41] Elisabeth Gössmann, "*Ipsa enim quasi domus sapientiae*: The Philosophical Anthropology of Hildegard von Bingen," *Mystics Quarterly* 13, no. 3 (1987): 146–54.

[42] See Heinrich Schipperges, "Ein unveröffentlichtes Hildegard-Fragment (Codex Berolin. Lat. Qu. 674)," *Sudhoffs Archiv für Geschichte der Medizin und der Naturwissenschaften* 40, no. 1 (1956): 41–77; Dronke believed in the passage's authenticity, see Peter Dronke, "Problemata Hildegardiana," *Mittellateinisches Jahrbuch* 16 (1981): 97–131.

In light of such problems, the contribution of Hildegard's secretaries and their potential influence on her literary output has remained subject to debate. Even though Hildegard warned her assistants not to interfere with the sense of her visions, and to restrict their corrections to formal aspects of the language,[43] the last decades of scholarship have seen palaeographical and computational evidence to maintain that the influence of these secretaries may as yet be an underexplored aspect of Hildegard's authorship. In chapter 6, we specifically focus on the role which Hildegard's secretaries played in shaping her image and authority, both during and after her life, by analyzing her (auto-)biography, the *Vita Hildegardis*. This dossier was intensely prepared in the final years of her life, both by herself and her secretaries, and has presented a source of fascination to scholars not only because it gives an account of Hildegard's life, but also because it contains direct citations of 'autobiographical' fragments allegedly written by the visionary herself.

### Suger of Saint-Denis

Suger (1080/1–1151) was a young man of fairly modest descent born just north of Paris. Around the age of ten, he was given as an oblate to the Benedictine abbey of Saint-Denis, close by his home village Chennevières-lès-Louvres.[44] Very soon it turned out this young man had a nose for ingratiating himself with the powerful. After years of loyal assistance to abbot Adam, Suger succeeded him in 1122, at which time he was forty-one/forty-two years old. He became counsellor of both Capetian kings Louis VI the Fat (1081–1137) and his son Louis VII († 1180), and also wrote *vitae* for both father and son. Near the end of his life, Suger would even act as regent of France when Louis VII left for the Second Crusade in 1147–9. Suger's life story, in other words, is that of a strong and charismatic personality who seems to have been the architect of his own fate, and climbed up the social ladder to become both an influential figure of the Church and temporary ruler of France.

The analogy of Suger as architect is not far-fetched: he literally *was* an architect. Suger spent a lifelong career not only constructing his own image, but also that of Saint-Denis. He enhanced the appeal of the abbey by renovating its basilica, which is the very first manifestation of high medieval Gothic architecture, and transformed it in one of the most popular pilgrimage destinations of the twelfth century. The abbey was prestigious and wealthy, and had ever since Merovingian times placed high stakes in presenting itself as fostering close ties to consecutive French dynasties. The Sandionysian patron saint, Dionysius the Areopagite, would in the early twelfth century come to be depicted on the banner of Vexin, a banner which king Louis VI himself

[43] Joan Ferrante, "*Scribe quae vides et audis*. Hildegard, Her Language, and Her Secretaries," in *The Tongue of the Fathers: Gender and Ideology in Twelfth-Century Latin*, ed. Townsend David, Taylor Andrew, and Ruth Mazo Karras, The Middle Ages Series (Philadelphia: University of Pennsylvania Press, 1998), 103.

[44] John F. Benton, "Introduction: Suger's Life and Personality," in *Abbot Suger and Saint-Denis. A Symposium*, ed. Paula Lieber Gerson (New York: The Metropolitan Museum of Art, 1986), 3.

held in fief from Saint-Denis since August 1124. This act made even the king of France appear as a vassal to the royal abbey.

Suger did not only write history in a figurative sense: he expressly instigated the composition of royal historiography in the scriptorium of Saint-Denis. Given these circumstances, Suger's sphere of influence extends somewhat more into an administrative and diplomatic milieu, and presents us with an abbot deeply ingratiated with secular powers. As no other, Suger understood the political power of language, and simultaneously —being the architect and parvenu that he was— he understood that whatever ties were non-existent could well be forged. In chapter 7, we assess the authorship of a suspicious Latin charter, titled the 'donation of Charlemagne,' which has in the past been brought in relation to Suger and his chancery due to the claims it makes on Saint-Denis' close ties with French royalty and its privileged position as "head of all the kingdom's churches" (*caput omnium ecclesiarum regni nostri*).

### Heloise and Abelard

The love story of Heloise and Abelard is one of the best known and most romanticized of the Middle Ages. Peter Abelard (1079–1142) was a flamboyant, reputed writer and dialectician, who reached the peak of his fame as a teacher in Paris c. 1116/7. At this time he began an amorous affair with Heloise of Argenteuil, who was his dearest and most talented pupil. When Heloise's uncle, Fulbert, discovered to his dismay that the couple's activities extended beyond those intellectual, it gave rise to a series of heated conflicts with Abelard, which ultimately culminated in the latter's brutal, nocturnal castration by a band of Fulbert's henchmen. This tragic event separated Heloise and Abelard (c. 1118/9), and made both of them convert to monasticism.

Current scholarship hypothesizes —and disagreement is still ongoing— that the lovers may be held responsible for having collaboratively composed two letter collections. The first is their conventionally 'accepted' letter collection consisting of eight letters. This collection is regarded to be one of the finest and most brilliant letter exchanges of the Latin Middle Ages. It constitutes a dialogue between Heloise and Abelard which first revolves around their history and relationship and gradually transforms into a foundational document for the Paraclete, the monastery founded by Abelard in 1122 and seven years later presided over by Heloise as prioress. The second collection is a set of 113 short letters exchanged between two anonymous lovers (<*V*>*ir* and <*M*>*ulier*), known as the *Epistolae duorum amantium*. Both of these collections have been subjected to fierce scholarly debates when it comes to their provenance, dating, intention and authorship. In chapter 8, we analyze them by applying stylometric methods.

# The Corpus

This is a thesis on 'distant'[45] as opposed to 'close' reading, which means that I will be statistically analyzing twelfth-century Latin texts by the hundreds instead of closely interpreting merely a handful of texts by their contents. It is impossible to continually make reference to the great number of texts analyzed per chapter, wherefore I will elaborate here on the distinctions between main corpus and benchmark corpus, categories which one will find mentioned throughout this thesis. Before doing this, I first provide a word on how the texts were provided in electronic format.

## Databases, Digitization and Disclosure

Aside from a number of exceptions, the pool of texts used for the experiments carried out in this thesis roughly derive from two main databases. The first is the Brepols Library of Latin Texts (LLT).[46] Especially the 'main corpus' as discussed below will be for a large part found collected in the LLT, generously provided by Brepols Publishers, who were partner to the research project "Collaborative Authorship in Twelfth-Century Latin Literature: A Stylometric Approach to Gender, Synergy and Authority."[47] The LLT first and foremost contains all the editions from Brepols's own *Corpus Christianorum*, in addition to several external critical editions that comply to modern critical standards.

Aside from being able to rely on the texts that had previously been collected in LLT, I have been fortunate to collaborate with Brepols on the further digitization of texts which had thusfar not been available in electronic edition, such as Könsgen's edition of the *Epistolae duorum amantium*,[48] Suger's collected oeuvre as edited by Gasparri,[49] or the respective works by Elisabeth and Ekbert of Schönau in the edition by Roth.[50]

For the remaining texts, I chiefly relied on the digitized version of the *Patrologia Latina* (*PL*), which has become electronically available since 1993, and has remained one of the most sizeable Latin corpora online (±113 million words).[51] The *PL* is a

[45] A term coined by Franco Moretti. On the distinction between close and distant reading, see p. 56.

[46] `http://www.brepolis.net`.

[47] This is the research project of which the current thesis is the result. The project was funded by the Ghent University Special Research Fund (BOF). Its execution rests on a close collaboration between the Henri Pirenne Institute for Medieval Studies at Ghent University, the CLiPS Computational Linguistics Group at the University of Antwerp, and the *Centre Traditio Litterarum Occidentalium* division for computer-assisted research into Latin language and literature housed in the *Corpus Christianorum* Library and Knowledge Centre of Brepols Publishers in Turnhout (Belgium).

[48] Ewald Könsgen, ed., *Epistolae duorum amantium. Briefe Abaelards und Heloises?*, Mittellateinische Studien und Texte 8 (Leiden and Cologne: Brill, 1974).

[49] Sugerius Sancti Dionysii, *Suger. Oeuvres*, ed. and trans. Françoise Gasparri, Les classiques de l'histoire de France au Moyen Age 37–41 (Paris: Les Belles Lettres, 1996–2001).

[50] Elisabeth Schoenaugiensis, *Die Visionen der hl. Elisabeth und die Schriften der Aebte Ekbert und Emecho von Schönau*, ed. Ferdinand W.E. Roth (Brünn: Raigerner Benediktiner Buchdruckerei. Verlag der "Studien aus dem Benediktiner- und Cistercienser-Orden", 1884).

[51] There is also a Greek counterpart publication, the *Patrologia Series Graeca* (161 volumes), which also includes Latin translations.

corpus containing texts of Latin ecclesiastical writers in 221 volumes ranging a time span of ten centuries, from Late Antiquity to the High Middle Ages (Tertullian c. 200 to Pope Innocent III c. 1216). The *PL* was first published in two series halfway the nineteenth century by the Parisian priest and theologian Jacques-Paul Migne (1800-1875), who mainly drew on seventeenth and eighteenth-century prints to compile the patristic heritage.

Migne's purpose was clear: "[...] I believe that I will grant the Church the greatest service it has ever been granted, and I well hope to die as the priest that has been its chief benefactor in the entire world by resuscitating its tradition integrally."[52] In this regard, Migne's work is both a blessing and a curse. His purpose was to be productive in a high pace, and in breadth and coverage he certainly succeeded (even though only half of his envisioned project attained completion).[53] On the other hand, the emphasis on quantity has obstructed its quality. Although Migne's conservation of a great number of patristic texts which would otherwise have gone lost unmistakably stands as a major contribution to medieval studies, his critical attitudes have been deficient, which has from time to time caused concerns over the corpus's overall reliability in matters of authenticity and ascription. However, the *PL*'s accessibility and encyclopedic quality render it a useful tool for scholars, and some texts can be consulted there and only there.

Therefore, the larger bulk of important texts in this thesis were not taken from the *PL*, but based on qualitative editions collected in the LLT, whose first and foremost advantage is that they are recent and critical publications which were carefully scrutinized by experts in the field. We will take care that at those instances where troublesome distortion by editorial principles should be located, they are presented as openly as possible.

Medieval texts have different orthographical appearances, and editors of texts in the LLT or those in the *PL* apply different rules and practices in transcribing texts and of handling and displaying the various witnesses. It stands beyond question that such differences constitute a poor ground upon which to automatically compare texts on a large scale, which is why they need to be addressed prior to proceeding to stylistic analysis. In natural language processing, this task commonly falls under 'preprocessing,' which entails minor interventions in the text such as the deletion of irrelevant textual material and the normalization of divergent orthographical forms, the principles of which will be explained more elaborately in chapter 3 and in the appendix.[54] In a nutshell, preprocessing Latin texts enables us to automatically align orthographical

[52] "[...] Je crois rendre à l'Église le plus grand service qui lui ait été rendu, et j'espère bien mourir le prêtre du monde entier qui lui aura fait le plus de bien, en ressuscitant intégralement sa tradition." I quote from a transcription of a letter written by Migne on 13 February 1858, published in Adalbert-Gautier Hamman, "Le centenaire de Jacques Paul Migne (1800–1875)," *Sacris Erudiri* 22 (1974): 13. Hamman provides no further details on how to find the letter except for its being located in the Bibliothèque Nationale (BN).

[53] Ibid., 13.

[54] See p. 90 and p. 313.

differences between such variant appearances of lexical items such as the pairs *racio* and *ratio*, or *aliquandiu* and *aliquamdiu*. Failing to attend to such differences, no matter how small as they may appear to the human eye, may in fact lead to disastrous misinterpretations of the results.

The text data used in this thesis will be open for consultation online in a GitHub repository,[55] together with accompanying code in Python,[56] which allows for replication of the experiments to be conducted on the pages to follow. Yet, the original texts have been slightly camouflaged so as to respect the copyright laws protecting the editions. Only function words —mentioned earlier as highly successful for distinguishing writing styles— were retained in their original position and form. All the remaining, content-loaded words, were substituted by 'dummy words', rendering the text illegible. This means that some experiments in this thesis, those which relied on most-frequent content words in addition to function words, will not be replicable by relying solely on the text data as available on GitHub. To replicate these experiments, one may request access to the electronic versions of the editions referred to by contacting Brepols.

## Main Corpus

The main corpus are those primary sources (cited from p. 377 onward) used for statistical analysis. In other words, this is a group of texts which does not only contain texts or passages which I have read and interpreted on a qualitative level, but have been subjected to quantitative analysis as well. One can think of Bernard's letter corpus in chapter 4, Elisabeth's visionary diaries in chapter 5, Hildegard's *Vita* in chapter 6, Suger's complete oeuvre in chapter 7, and Heloise and Abelard's letter collection in chapter 8. For each chapter the exact versions and editions used for these texts will be recoverable in the footnotes, and therefore also in the primary sources at the back of this thesis. There is no separate index provided which enlists the entire main corpus: each chapter will make clear the provenance of these important texts per case study. Since the main corpus is composed of the most important texts of this thesis, they were often those works as provided from the LLT (mentioned on p. 18).

## Benchmark Corpus

The texts contained in the benchmark corpus are listed in the appendix (pp. 304–308.), and will be predominantly involved indirectly without becoming subject of closer study. The bulk of texts in this corpus largely stems from the *PL*, although some texts by authors Peter the Venerable, Rupert of Deutz, Honorius Augustodunensis, Hugh of Saint-Victor and Guibert of Nogent have been generously provided by Brepols

[55] `https://github.com/jedgusse/collaborative-authorship`

[56] A portion of the appendix is dedicated to programming, see pp. 309ff.

from the LLT (see p. 18 above). The benefit of the benchmark corpus is that it enabled to run a few trial experiments (especially in chapters 2 and 3) to fine-tune the methodology prior to beginning the case studies in part 2.

# Part 1

# 1

# Collaborative Authorship

## 1.1 From Scriptorium to Text

Since the selected case studies in this thesis explore collaborative authorship according to a thematic sliding from the reality of the scriptorium to the reality of the text, this first chapter imitates this larger structure. Firstly, we discuss the education, training and professional activities of secretaries, and their participation in the dictation, redaction and transcription of texts. This serves to emphasize the oral-aural origin of medieval literature and its circulation through many hands, typical to a manuscript culture, whose conditions automatically implied a higher degree of fluctuation and lack of authorial control. Simultaneously, we show that despite these mutable circumstances, twelfth-century authors fostered proper practices and customs in composing their texts, and exercised a personal authority and charisma that leaves open to question to what extent this mutability common to all medieval literature was restricted or enhanced in individual cases.

Of vital importance in this regard is the extent by which the textual contribution of all authors involved hinged on hierarchical structures exerted in the scriptorium and/or inscribed in medieval culture. In the course of this chapter I give a number of examples to illustrate that there were literary-social norms and values which demarcated the respective spheres of influence in writing collaborations, and cultivated distinctions between actions acceptable from transgressive in the process of composition. Writing was considered to be an expression of authority. Appropriation and attribution provided instruments of authorization and repression, wherefore secretaries operated in an atmosphere of utmost confidentiality.

A particularly interesting manifestation of the sensitive power dynamics at stake in medieval collaborative composition are the writing partnerships of female writers with male spirituals, provosts and/or confessors. Since women were discouraged from education in the schools, from taking up didactic positions, and of entertaining literary ambitions, they often relied on the endorsement of male supporters. This kind of cross-sex collaboration constitutes a unique field of force. On the one hand, female

writers such as Elisabeth of Schönau, Hildegard of Bingen and Heloise of Argenteuil devalued their proper authority in writing, and called on the assistance of men, but especially with visionaries Elisabeth and Hildegard this same act of self-devaluation simultaneously established a position from which they were able to assert a specifically female type of non-institutionalized learning, which was divinely inspired and based on charisma.

Finally, this chapter illustrates how collaborative authorship is to be understood in a broader sense than merely by its physical and material manifestations. The twelfth-century author was also in continual 'collaboration' with a tradition and a past, especially the Latin past. In such a culture, impersonation, text re-use, anthologization and compilation are an integral part of literary composition. Understanding the exact intention behind some of these 'adoptions' of text written by other authors is far from straightforward, especially when these forms of writing start to resemble what modern readership would commonly categorize under such terms as plagiarism, forgery or fictionality. The aptitude and ineptitude of these terms come to stand in an interesting tension when considering that the current thesis's methodology is heavily indebted to methods applied in modern-day plagiarism detection.

## 1.2 The Profession of Secretary

### 1.2.1 Secretaries in Premodern Times

Collaborative authorship was already common practice in the Graeco-Roman world of Antiquity —especially from the first century BC onwards—[1] and later. One can think of well-known examples of, for instance, Paul's (AD 5–c. 64/67) secretary Timothy (and others),[2] or Cicero's assistant Tiro (103–4 BC), who was first slave, then freedman, and counts as the inventor of the earliest of shorthand systems in Latin, the *notae Tironianae*.[3] Importantly, secretaries' roles in a premodern textual culture extended towards an everyday practical and societal function.[4] Although the textual cultures of Antiquity and the Middle Ages should not be equated too lightly, the phenomenon of authors dictating to a secretary seems to resonate most predominantly in bureaucratic or 'bureaucraticizing' cultures. In other words, these are cultures in which the written word gains an alternative function —or even gains supremacy— as opposed to orally delivered text (although the two forms of writing continued to have

[1] Anthony Di Renzo, "His Master's Voice: Tiro and the Rise of the Roman Secretarial Class," *Journal of Technical Writing and Communication* 30, no. 2 (2000): 158.

[2] Ernest Randolph Richards, *The Secretary in the Letters of Paul*, Wissenschaftliche Untersuchungen zum Neuen Testament 2:42 (Tübingen: Mohr, 1991).

[3] Di Renzo, "His Master's Voice," 156.

[4] For especially Christian Late Antiquity, and an attempt to an exhaustive classification of secretarial roles in this period, see Kim Haines-Eitzen, *Guardians of Letters. Literacy, Power, and the Transmitters of Early Christian Literature* (New York: Oxford University Press, 2000).

very close interaction and cannot be strictly separated from one another). Secretaries were not merely a part of a creative collaboration in the composition of speeches and literature, but they had very pragmatic and supportive roles in the everyday life of the ruling elite, in judicial and legal instances or in the trades administration of businessmen and merchants.

### 1.2.2 The *Artes*

The medieval author's encirclement by secretaries meant the formation of a bastion of scribes, copyists and assistants around a *dictator*, and —as automatically would have been the case in a physical manuscript culture—, a library or collection of authoritative writings. Secretaries were specifically trained to imitate their *dictator*. To an extent these were distinctive schools burgeoning around the influence of a single figure, whose name —and style— would come to grant the work its authority. For some of these teams —such as those encircling Bernard of Clairvaux or Peter the Venerable (c. 1092–1156)— one might be inclined to contend that they were chanceries.[5] This relates the phenomenon of trained imitators to what Giles Constable termed the "birth of modern bureaucracy."[6] That is not to say that the teams surrounding these authors would constitute a formal or institutional body of administration charged with the composition and dispatch of official documents as such. What they do testify of, on the other hand, is associations with a more general movement of increased professionalization and bureaucratization of writing, which in its own turn stands in conjunction with the twelfth-century transformations of the educational system.

The most patent example of how Latin education was increasingly oriented towards favourable career prospects, is the rise of the discipline of *ars dictaminis*, alternatively encountered as *dictamen* and *ars dictandi*.[7] In technical terms its rise is indicated by the arrival of influential theoretical manuals and ordinances which regulated the writing style of letters and decrees, or the 'art of dictation,' of which the earliest examples are Alberic of Monte Cassino's *Flores rhetorici* (or *Dictaminum radii*) and *Breviarium de dictamine* (fl. 1079). Broadly, these handbooks contained rhetorical guidelines on Latin style for both prose and poetry, with a dominant focus on epistolography.[8] Who exactly invented the *ars dictaminis* first, and in what way its precepts are to be

[5] Rassow went as far as to equate the workings in Bernard's scriptorium with a chancery, see Peter Rassow, "Die Kanzlei Bernhards von Clairvaux," *Studien und Mitteilungen zur Geschichte des Benediktiner-Ordens* 34 (1913): 243–79.

[6] Giles Constable, "Dictators and Diplomats in the Eleventh and Twelfth Centuries: Medieval Epistolography and the Birth of Modern Bureaucracy," *Dumbarton Oaks Papers* 46 (1992): 37–46.

[7] The bibliography on the *ars dictaminis* is extensive. For further reference I cite the important works of Giles Constable, *Letters and Letter-Collections*, Typologie des sources du moyen âge occidental 17 (Turnhout: Brepols, 1976); William D. Patt, "The Early 'Ars Dictaminis' as Response to a Changing Society," *Viator* 9 (1978): 133–56; Ronald Witt, "Medieval 'Ars Dictaminis' and the Beginnings of Humanism: a New Construction of the Problem," *Renaissance Quarterly* 35, no. 1 (1982): 1–35.

[8] See Patt, "Early 'Ars Dictaminis'," 133–4.

connected to the art of letter-writing in (Late) Antiquity, is a matter of dispute.[9] In any case it is clear that the *artes* knew an unprecedented growth in Italy during the final decades of the eleventh century, before they found their way North and became of great importance in the curricula of the schools. The *ars dictaminis*' main focus lay on the form, rhythm and arrangement of Latin prose, and the application of certain formulas. Diplomats, ambassadors, and secretaries would inspire one another in a network of correspondence, or share suchlike rhetorical devices within their scriptoria. A figure such as Nicholas of Montiéramey († 1176/78), one of Bernard of Clairvaux's most important secretaries, would have met many of the demands that were set to the chancery clerks of the day.[10]

## 1.3 The *Reportatio*

### 1.3.1 Dictation

A writing collaboration between the author and his or her secretary would have roughly proceeded as follows.

1. A text —or blueprint of a text— was devised within an oral setting.
2. The rough outlines of the declamation were sketched on a wax tablet with a *stylus* by either the author or a secretary.
3. The rudimentary text was transcribed from wax to parchment, generally by the secretary.

One could argue that medieval literature was transmitted through three different carriers: the mouth, the wax tablet, and parchment. This is a simplified schema, of course, which knew many variants dependent on the distinctive case at hand. Generally, texts would have known their inception in an oral-aural interaction between a dictator and a secretary, as had been common practice in the preceding centuries.[11] A sermon or letter was spoken aloud, or perhaps delivered in front of an audience.

[9] See especially Patt, "Early 'Ars Dictaminis'."

[10] Anne-Marie Turcan-Verkerk, "L'Introduction de l'*Ars Dictaminis* en France. Nicolas de Montiéramey, un professionel du dictamen entre 1140 et 1158," in *Le dictamen dans tous ses états. Perspectives de recherche sur la théorie et la pratique de l'ars dictaminis (xi^e^-xv^e^ siècles)*, ed. Benoît Grévin and Anne-Marie Turcan-Verkerk (Turnhout: Brepols, 2015), 63–98.

[11] Saenger has, however, convincingly shown that this oral reading culture was accompanied by a tendency of silent copying, which indicates the gradual replacement of oral forms of Roman dictation and self-dictation. Saenger's main aim was to contradict that silent reading arrived after the invention of the printing press in the late fifteenth century. Transitions from an oral to a silent reading culture are reflected in monastic manuscripts from Late Antiquity and the Early Middle Ages, which increasingly show word spacing. The spacing would emphasize the increased importance of visual structures on parchment, allowing scribes to interpret and read in silence instead of having to pronounce the words to themselves. See Paul Saenger, "Silent reading: Its Impact on Late Medieval Script and Society," *Viator* 13 (1982): 379; for the twelfth century specifically, see the fourteenth chapter "Reading and Writing in Northern Europe in the Twelfth and Thirteenth Centuries," in Saenger, *Space Between Words*, 243–55.

Consequently, once the text had gained a rudimentary form in an oral phase, it was engraved on a wax tablet with a *stylus* in a kind of shorthand called tachygraphy.[12] A wax tablet (*tabula*, *tabella cerata*, *cerae*, ...) consisted of two or more wooden boards sewn together with the inside surfaces hollowed out and coated in wax. The *stylus* was a sharp writing tool, made of wood, metal or bone, by which to carve into the wax, which could also be heated in order to be more malleable.[13] The wax tablet and the *stylus* were the first and elementary tools for writing down what had been spoken. Its superiority as opposed to parchment scraps or single sheets —which could indeed be used for drafting as well— was constituted by the fact that wax tablets were reusable and more easy to transport. They would have been a part of authors' basic education in Antiquity until well into the sixteenth century.

This strong connection between wax tablets and learning to speak and write can be found emphasized in the depicted personifications of *Rhetorica* —one of the liberal arts— who is often found holding a wax tablet.[14] Isidore of Seville (560–636) would come to see the wax metaphorically as the author's mind that could be kneaded or engraved with knowledge. He referred to it as the "matter of letters, the 'nurse' of children, granting young boys intellect and the beginnings of reason."[15] And yet, it is important to keep in mind that the material and the technical was only part of what constituted literacy and literary composition, and that wax tablets were regarded as tools by which literature was memorized. One could refer to the much debated passage in which Einhard (c. 770–840), Charlemagne's secretary and biographer, narrates that the emperor used to sleep with wax tablets underneath this pillow, so that whenever the emperor had his hands free he could learn to write.[16] The biographer suggests that Charlemagne had learned to do so only too late in life, and therefore never truly got the hang of it. Charlemagne's inability to get a full grasp on the technical aspects of writing has been taken by some to denote that the emperor had been illiterate, but to assert this must also partly be an anachronistic understanding of what constituted medieval literacy.[17] Charlemagne was simultaneously deemed to be a great orator,

[12] For the usage of wax tablets, and their importance in the medieval education of letters, see Richard H. Rouse and Mary A. Rouse, "Wax Tablets," *Language and Communication* 9, nos. 2/3 (1989): 175–91; for tachygraphy, see Malcolm Beckwith Parkes, "Tachygraphy in the Middle Ages: Writing Techniques Employed for *Reportationes* of Lectures and Sermons," in *Scribes, Scripts and Readers: Studies in the Communication, Presentation and Dissemination of Medieval Texts* (London: Hambledon Press, 1991), 19–33.

[13] For a comprehensive survey of testimonies of authors on using wax tablets, see Wilhelm Wattenbach, *Das Schrifwesen im Mittelalter*, 3rd ed. (Leipzig: Verlag von S. Hirzel, 1896 (1871)), 51–89.

[14] Rouse and Rouse, "Wax Tablets," 176.

[15] My translation. "Cerae litterarum materies, parvulorum nutrices, ipsae 'Dant ingenium pueris primordia sensus,'"in Isidorus Hispalensis, *Etymologiarum libri XX*, in *PL* 82:73–718, ed. Faustinus Arevalus, see 239; discussed in Rouse and Rouse, "Wax Tablets," 176.

[16] "Temptabat et scribere tabulasque et codicellos ad hoc in lecto sub cervicalibus circumferre solebat, ut, cum vacuum tempus esset, manum litteris effigiendis adsuesceret, sed parum successit labor praeposterus ac sero inchoatus," see Einhardus, *Einhard. Vita Karoli Magni*, ed. Ludwig Gompf (Münster: Verlag Aschendorff, 1971), § 25, ll. 26–30, 25.

[17] On a broad discussion of the passage, see chap. 3 titled "Karolus magnus scriptor"in Paul Edward Dutton, *Charlemagne's Mustache and Other Cultural Clusters of a Dark Age*, The New Middle Ages (New York: Palgrave, 2004),

both in Latin as in his mother tongue.[18] At least to Einhard, Charlemagne's inabilities in physically forming letters on a wax tablet does not univocally coincide with the inability to compose literature.

Indeed, nearly all compositions of medieval literary works were firmly rooted in an interaction between the spoken and written word. In some contexts, such as that of Bernard of Clairvaux, the concurrence of these two modes of writing is reflected in how *dictare* had come to supplant the meaning of *scribere* to designate the act of authorship.[19] Dictation to a secretary simultaneously implied that record was taken for future transcription. Presumably, a preacher such as Bernard would also have practiced his sermons aloud, or would have delivered them in their rudimentary shape to a test audience (such as the monks of Clairvaux).[20] The consequent recording of the text was a long and intricate process. Either Bernard himself drafted the rough structures on a wax tablet, or else he orally dictated cues, keywords, biblical references and a rough outline of their disposition to one or more secretaries. Once this fundamental transference from sound to sign had taken place, secretaries copied the text from the wax tablets (*excipere*) to parchment (the *membrana*, or animal 'skin').

### 1.3.2 Autographs

What the exact procedures of the *reportatio* were, and in what way the first drafting of texts on wax tablets was carried out, is difficult to reconstruct. Additionally, one can hardly expect that in every situation and for every type of text a similar procedure would have been maintained. The text's conception in an oral setting and its intermediary transference to effaceable and often irrecoverable carriers heavily complicates the question if what lies before us is still the work of the 'original' author. We can assume that the material conditions automatically stipulated that a text travelled some time before reaching a destination, changing its guises along the way. Yet, this trajectory can rarely be ascertained. We can only rely on second-hand sources (not always very reliable ones) such as miniatures or written records, that give us some impression. In two medieval miniatures featuring Hildegard of Bingen and her secretary Volmar of Disibodenberg (fig. 1.1), whose design she might have been responsible for

69–92.

[18] "Erat eloquentia copiosus et exuberans poteratque, quicquid vellet, apertissime exprimere. Nec patrio tantum sermone contentus, etiam peregrinis linguis ediscendis operam impendit. In quibus Latinam ita didicit, ut aeque illa ac patria lingua orare sit solitus, Graecam vero melius intellegere quam pronuntiare poterat," see Einhardus, *Vita Karoli Magni*, § 25, ll. 9–15, 25.

[19] Ernst Robert Curtius, *European Literature and the Latin Middle Ages*, trans. Willard R. Trask, Bollingen Series XXXVI (Princeton: Princeton University Press, 2013 (1948)), 163; Saenger, "Silent Reading," 380–1.

[20] The workings of Clairvaux's scriptorium are extensively investigated in Rassow, "Die Kanzlei Bernhards von Clairvaux"; for the tension between the spoken word and the recorded text referred to, see Giles Constable, "The Language of Preaching in the Twelfth Century," *Viator* 25 (1994): 134–5.

(a) c. 1165. MS Wiesbaden, Hessische Landesbibliothek, 1, fo. 1[r]. The *Scivias* manuscript was lost after World War II. This reproduction dates from 1927-1933.[22]

(b) c. 1220-1230. MS Lucca, Biblioteca Statale, 1942, fo. 1[v]. One of ten illuminations.

Figure 1.1: Miniatures of Hildegard of Bingen receiving inspiration from the Holy Ghost (the red, fiery flame), and engraving her texts on a wax tablet. She is sided by her secretary Volmar of Disibodenberg, and in the right-hand figure also by a female scribe (perhaps Richardis of Stade). There is debate as to whether or not Hildegard herself would have designed the illuminations.

herself,[21] we observe that it is she who engraves the wax tablet with the *stylus*, and her collaborator who notes down words on parchment.

However, whether or not these depictions are representative for the historical reality of the composition process, is difficult to answer. In both miniatures, Hildegard appears to be writing directly onto the wax tablets, while simultaneously being heard and recorded by her secretary from the other side of the enclosure. As Albert Derolez has argued, the suggestion that Hildegard is dictating and writing simultaneously seems absurd.[23] The miniature appears to have an agenda of amplifying Hildegard as author

[21] This hypothesis is fully described in Madeline H. Caviness, "Hildegard as Designer of the Illustrations to Her Works," in *Hildegard of Bingen: The Context of Her Thought and Art*, ed. Charles S.F. Burnett and Peter Dronke, Warburg Institute Colloquia 4 (London: Warburg Institute, 1998), 29–62. Counter-arguments have been formulated in Lieselotte Saurma-Jeltsch, "Die Rupertsberger *Scivias*-Handschrift. Überlegungen zu ihrer Entstehung," in *Hildegard von Bingen. Prophetin durch die Zeiten. Zum 900. Geburtstag*, ed. Edeltraud Forster (Basel/Freiburg im Breisgau/Vienna, 1997), 340–58.

[22] Gottfried Zedler, *Die Handschriften der Nassauischen Landesbibliothek zu Wiesbaden*, Zentralblatt für Bibliothekswesen 63 (Leipzig: Harrassowitz, 1931).

[23] Albert Derolez, "Deux notes concernant Hildegarde de Bingen," *Scriptorium* 27 (1973): 292.

of her own works, and her secretary as a mere corrector. On the other hand, the female figure standing besides Hildegard on the right-hand Lucca miniature seems to be represented passively, which may in its own turn designate the different levels of authority between male and female scribes. Whether or not Hildegard would dictate instead of write her texts, and whether or not Volmar's role was restricted to slight amendations on the level of style, or to large interferences on his own behalf, is only the question.

This tension between author and secretary raises all kinds of intricate questions concerning Hildegard's professed deficiencies in the writing and reading of Latin, and to her collaboration with male clerics, to which we will come back later in this chapter. What deserves emphasis here, is that already at the earliest stages of medieval authorship, where the dictated text was transferred to wax tablets, there is an ambiguity of roles. Moreover, the signs and abbreviations on wax tablets must have come closest to any kind of authentic, individual expression —or autograph— by a medieval author. For almost every written record of the Middle Ages, the absence of autographs constitutes the 'missing link' in reconstructing the author's original speech and style. Once a scribe became involved, medieval authors could —often to their own frustration— see their work being tampered with. It is therefore striking that from the thirteenth century onwards a comparably larger number of depictions of authors writing with their own hand —in complete solitude— can be encountered. Wax tablets would increasingly become obsolete, and authors would more often write their own autographs, which would have been far less common in the twelfth century.[24]

### 1.3.3 Bound to the Material

In most cases, a fair copy of a text on parchment was made only as soon as the composition was in its entirety engraved on wax tablets. Practically, this meant that a considerable number of wax tablets was prepared before transcription began. This can be deduced, for example, from Anselm of Canterbury's (1033/4–1109) *Vita*, in a passage where his biographer and secretary Eadmer recounts that the wax tablets on which Anselm had written his *Proslogion* were first lost, then shattered, leaving them with no secondary copy of the text.[25] The loss or lack of wax tablets considerably endangered the literary undertaking. Much of the preparation, organization and economical resources depended upon material factors, which makes that medieval authorship indeed has a very close connection to physical carriers. To a great extent, the limits of the author were constituted by the limits of his material. In many ways,

[24] Saenger, *Space Between Words*, 257; for autographs (and specifically the lack thereof) in the eleventh and twelfth centuries, see Monique-Cécile Garand, "Auteurs latins et autographes des XI$^{e}$ et XII$^{e}$ siècles," *Scrittura e civiltà* 5 (1981): 77–108.

[25] The destruction of the work would have been the work of the devil. See Eadmerus Cantuariensis, *The Life of St. Anselm, Archbishop of Canterbury*, ed. and trans. Richard William Southern, Medieval Texts (London: Nelson, 1962), I.xix, 30–1.

one could argue that scribes were an instrumental extension of the author's capacity to write, or at least, if they saw eye to eye.

In Baudri of Bourgeuil's poems (c. 1050–1130) addressed to his scribe(s), we catch a glimpse of how this could go wrong. Wax tablets could only provide a limited amount of space, and scribes were therefore expected to keep a pace. Abbot Baudri complains of the lazy scribe Gerard, whose transcriptions of Baudri's poems to parchment are too slow. Gerard cannot keep up with Baudri's unbridled creative outpouring, wherefore the latter is forced to write poem on top of poem in his wax tablets.[26]

> As thus, my dear Gerard, may you go from one foot like another,
> As all that I desire from you is to write my poems.
> I would have written poems on top of my poems,
> If the tablets could hold any more.
> I have filled my tablets, while you were idling,
> while you were slow to transcribe what is in the wax.
> But so that you may empty the wax, transcribe my work ;
> so that virtuously you may shake off your laziness.[27]

Competent personnel is hard to get by! We also read of instances of overzealous scribes, who in all their enthusiasm fail to grasp the sense of the message that was meant to be conveyed. Bernard of Clairvaux's increased discomfort concerning the fact that others wrote in his name, and his dissatisfaction with some of them when it came to grasping the *sensus* of his message, becomes clear in a short apologetic letter addressed to Peter the Venerable. A monk named Nicholas (who might possibly be identified as the head librarian Nicholas of Montiéramey) had informed Bernard that he had been able to glance upon what was probably a copy of an outgoing letter addressed to Peter containing "words of bitterness" (*voces amaritudinis*). Bernard was quick to draft and send a new letter to the abbot of Cluny to guarantee that the letter had not been of his making, but that of a scribe:

> Believe me, I who am fond of you, that what has offended you did not originate in my heart, nor was it taken from my mouth. The excess of work is to blame, because while my scribes do not understand well what I mean, they sharpen their *stylus* beyond measure, and I cannot see what I have ordered them to write.[28]

[26] See the poem titled *Ad Girardum scriptorem suum*, edited in Baldricus Dolensis, *Poèmes*, ed. and trans. Jean-Yves Tilliette, Auteurs latins du moyen âge (Paris: Les Belles Lettres, 1998), 1:34.

[27] My translation. "Sic, Girarde meus, tibi sit pes unus ut alter, | Quodque tibi cupio, carmina scribe mea. | Carmina carminibus nostris superapposuissem, | Si superapposita susciperent tabulae. | Impleui nostras, dum tu pigritare, tabellas, | Dum scriptum in cera lentus es excipere. | Vt uero ceram uacues, opus excipe nostrum ; | Vt probus a solita te excute pigricia." The original Latin and a French translation is found in Tilliette's edition of Baudri's poems, already referred to in n. 26

[28] My translation. "Credite amanti, quia nec in corde meo ortum est, nec ab ore meo extortum est, quod aures vestrae beatitudinis exasperaret. Multitudo negotiorum in culpa est, quia dum scriptores nostri non bene retinent sensum nostrum, ultra modum acuunt stilum suum, nec videre possum quae scribi praecepi," see Ep. 387 in Bernardus Claraevallensis, *Sancti Bernardi Opera* (*SBO*), ed. Jean Leclercq, Henri M. Rochais, and Charles Holwell Talbot (Rome: Editiones Cistercienses, 1957–77), 8:355.

Due to the material conditions, medieval text was vulnerable to fluctuation. Scribes and by extension future copyists possessed over the potential to interfere with authors' works. Not only to Bernard, but also to Elisabeth of Schönau, this appeared to have been a concern when she stated that "certain people are circulating letters [...] in my name."[29] Elisabeth testifies of this in a letter to Hildegard of Bingen, her older visionary role model, who had undoubtedly experienced similar concerns. In the final phrases of her *Liber divinorum operum*, Hildegard had in her own turn warned the reader that whoever dared to alter her words would indulge in great sin, and enter in conflict with the greatest of possible authorities, namely the Holy Spirit:

> Whence let no man be so bold as to add anything to the words of this writing or take anything away, lest he be deleted from the book of life and from every blessedness that is beneath the sun, except for the correction of letters or words, which were put forth simply through the inspiration of the holy spirit. Whoever presumes otherwise sins against the holy spirit. Which sin will not be remitted neither here nor in the next world.[30]

Guibert, the later abbot of Nogent-sous-Coucy (c. 1055–1124), appears to have claimed in his *De vita sua* (alternatively *Monodiae*) that he wrote all of his texts straight to parchment, a habit which would have originated out of necessity.[31] Guibert's former abbot in Saint-Germer had distrusted the young monk's fondness for reading and writing, therefore prohibiting him to do so. A lucky coincidence nevertheless enabled Guibert to find parchment on which to write his commentary on Genesis. Consequentially, Guibert's authorship is one of the rare occasions where we find the writer practicing his métier in solitude and silence. In the preface to his *Dei gesta per Francos*, Guibert asks his reader to excuse his style, because he had not drafted the text on wax tablets before applying them to parchment.[32] Guibert

[29] "Audio, et quosdam litteras de suo spiritu scriptas sub nomine meo circumferre."See her *Liber tercius visionum*, in Elisabeth Schoenaugiensis, *Visionen*, §19, 71; translation taken from Elisabeth Schoenaugiensis, *Elisabeth of Schönau: The Complete Works*, trans. Anne L. Clark and Barbara Newman, The Classics of Western Spirituality (New York: Paulist Press, 2000), 138.

[30] "Unde nullus hominum tam audax sit, ut uerbis huius scripture aliquid augendo apponat uel minuendo auferat, ne de libro uite et de omni beatitudine que sub sole est deleatur; nisi propter excribrationem litterarum aut dictionum, que per inspirationem Spiritus Sancti simpliciter prolata sunt, fiat. Qui autem aliter presumpserit, in Spiritum Sanctum peccat. Unde nec hic neque in futuro seculo illi remittetur," see Hildegardis Bingensis, *Liber divinorum operum*, ed. Albert Derolez and Peter Dronke, *CC CM* 92 (Turnhout: Brepols, 1996), 3.5.38, 44–5, 462–3; translation taken from Ferrante, "*Scribe quae vides et audis*," 105.

[31] "Opuscula enim mea haec et alia nullis impressa tabulis, dictando et scribendo, scribendo etiam pariter commutando, immutabiliter paginis inferebam," in Guibertus de Novigento, *Autobiographie*, ed. and trans. Edmond-Réné Labande, Les classiques de l'histoire de France au Moyen Age 34 (Paris: Les Belles Lettres, 1981), 1.17, l. 5, 144; the passage is discussed in Wattenbach, *Das Schrifwesen im Mittelalter*, 71; and likewise in Monique-Cécile Garand, *Guibert de Nogent et ses secrétaires*, Autographa Medii Aeui II (Turnhout: Brepols, 1995); Mortensen has formulated doubts as to whether or not Guibert's assertion is to be taken all too seriously, emphasizing the common misreading of the phrase cited here. The word *commentando* in some manuscripts versions reads *commutando*, which changes the meaning of 'commending' texts to paper into 'exchanging' the text's medium. See Lars Boje Mortensen, "From Vernacular Interviews Into Prose (ca. 600–1200)," in *Oral Art Forms and Their Passage Into Writing*, ed. Else Mundal and Jonas Wellendorf (Copenhagen: Museum Tusculanum Press, 2008), 63, n. 26.

[32] "[...] nec ceris emendanda diligenter excepi sed uti presto est fede delatrata membranis apposui," in Guibertus de

also explicitly stated that he disliked working with secretaries, despite the fact that he must have had experience with them.[33] Definitely at the end of his life, when he became blind, he deplored the fact that he could no longer see his works with his own eyes, and revise his own style. In his preface to the *Tropologiae in Osee, Amos ac lamentationes Jeremiae*, we read the following passage:

> Not only through my perpetual reading but also through my continual writing —I have thusfar been not only dictator, but also notary of tireless labor— I had blunted the sharpness of my eyes completely. My immoderation in this reduced me to the extent that I appointed a scribe, which had never been my custom, and the present work forced me to explain everything from memory, by voice, without a hand, and without eyes. To this cause, may the moderation of a virtuous reader attend, that if he were to find me speaking in unusual mode or saying something less competently, that by this insight he does not hesitate in indulging me, but less so the other hand writing down what is mine and which —so he may recognize— I had once been accustomed to putting down myself. For when I wrote my diction with my own hand, I often and continually revised those same words while I wrote, and it was much easier for me to insert what was missing. While I had nothing to fear of conflict arising from my notary's laziness, meticulous as I was I could caressly apply myself to the accuracy of my words which were my own affair.[34]

## 1.4 A Delicate Relationship

### 1.4.1 Bio-/Hagiography and Canonization

The examples above already suggest that the collaboration between author and secretary was a sensitive project, which came with a close and confidential bond. The author and the secretary would spend a considerable amount of time in each other's company, reading and writing intensively. A secretary will have been well aware of an institution's state of affairs, and the author's state of mind. In the end, (s)he was a close friend and associate, who was largely responsible in drawing up the image, memory and reputation of an author for posterity, and (in some cases) the commu-

Novigento, *Guibert de Nogent:* Dei gesta per Francos *et cinq autres textes*, ed. Robert B.C. Huygens, *CC CM* 127A (Turnhout: Brepols, 1996), in his *praefatio*, l. 118.

[33] On Guibert of Nogent and his collaboration with secretaries in the scriptorium, see Monique-Cécile Garand, "Le scriptorium de Guibert de Nogent," *Scriptorium* 31, no. 1 (1977): 3–29; and Garand, *Guibert de Nogent et ses secrétaires*, especially 26–9.

[34] My translation. "Hactenus enim non tam perpetuitate legendi quam nimia continuatione scribendi, utpote qui non solum dictator exstiteram, sed et laboris indefessi notarius, oculorum meorum aciem undecunque obtuderam, unde ad hoc mea immoderantia me redegit, ut exceptore adhibito, quod mihi nunquam moris fuerat, sola memoria, sola voce, sine manu, sine oculis praesens opusculum cogeret explicari. In quo pii moderatio lectoris attendat ut, si insolito mihi modo dictantem minus aliqua competenter dixisse repererit, eo mihi intuitu indulgere non differat, quo minus alienae dum mea scribit manui, quam meae quondam facere consueveram institisse cognoscit. Dum enim mea manu propria scriptitarem, et crebro contuitu inter scribendum eadem dicta reviserem, facillimum mihi erat et omissa retexere, et dum notarii mei fastidia nulla ex mora revereor, verborum curialitati secure mihimet ipsi morosus intendo," edited in Guibertus de Novigento, *Tropologiae in prophetas Osee et Amos ac lamentationes Jeremiae*, in *PL* 156:337–488, ed. Luc d'Achery, 340. I was able to partially support my English translation by aid of the French one in Garand, *Guibert de Nogent et ses secrétaires*, 27–8.

nity represented by that author. (S)he was —often both implicitly as explicitly— biographer (or hagiographer) of the author in question. This was the case for Eadmer and Anselm of Canterbury,[35] Suger of Saint-Denis and the French kings Louis VI and Louis VII (fragmentary),[36] William of Saint-Denis and that same Suger,[37] Geoffrey of Auxerre and Bernard of Clairvaux,[38] Ekbert and Elisabeth of Schönau,[39] Guibert of Gembloux and Hildegard of Bingen,[40] etc.

At times we have the loyalty of scribes to thank that any documentation on their *dictator*'s life is known to us at all. This is true for Eadmer, who ignored Anselm's explicit demand that his *Vita* were destroyed.[41] A secretary's responsibility and textual influence can be seen to extend well beyond the author's death. Ultimately, the documentation and bundling of the author's textual production into a sort of dossier could prove instrumental for the author's canonization. In the twelfth century, the procedures leading towards canonization were increasingly systematized and bureaucratic, meaning that the papal See's decision "rather than [...] popular veneration and

[35] Eadmerus Cantuariensis, *The Life of St. Anselm*.

[36] For the so-called *Vita Ludovici grossi Regis*, see Sugerius Sancti Dionysii, *Vie de Louis VI le gros*, 2nd ed., ed. and trans. Henri Waquet, Les classiques de l'histoire de France au Moyen Age 11 (Paris, 1964 (1929)); and for the fragmentary *De Glorioso Rege Ludovico, Ludovici Filio*, see Sugerius Sancti Dionysii, *Histoire de Louis VII*, in *Suger. Oeuvres*, ed. and trans. Françoise Gasparri, vol. 1, Les classiques de l'histoire de France au Moyen Age 37 (Paris: Les Belles Lettres, 1996), 156–77; it is well-known that Suger has functioned (at some occassions) as secretary for Louis VI, and the intervention of his Latin style has been witnessed in charters of the royal entourage from 1118 onwards. See Jean Dufour, ed., *Recueil des actes de Louis VI, roi de France (1108-1137)*, vol. 3, Chartes et diplômes relatifs à l'histoire de France (Paris: Diffusion de Bocard, 1993), 60.

[37] Guillelmus Sancti Dionysii, *Vie de Suger par le moine Guillaume*, in *Oeuvres / Suger*, ed. and trans. Françoise Gasparri, vol. 2, Les classiques de l'histoire de France au Moyen Age 41 (Paris: Les Belles Lettres, 2001), 292–357.

[38] See Guillelmus de Sancto Theodorico, Arnaldus Bonaevallensis, and Gaufridus Autissiodorensis, *Vita prima Sancti Bernardi Claraevallis abbatis*, in *Guillelmi a Sancto Theodorico opera omnia*, ed. Paul Verdeyen, vol. 6, *CC CM* 89B (Turnhout: Brepols, 2011), 13–233. The *Vita prima* is in fact the work of three authors, William of Saint-Thierry (the first book), Arnaud of Bonneval (the second) and Geoffrey of Auxerre (the three remaining books). The *Vita prima* exists in two versions. The first version, known as *Recensio A*, is edited in the aforementioned version and in *PL* 185:225–68. The initial case for canonization for Bernard was rejected by Innocent II. *Recensio B* is a further amendation by Geoffrey of Auxerre, and is edited (and translated) in Guillelmus de Sancto Theodorico, Arnaldus Bonaevallensis, and Gaufridus Autissiodorensis, *First Life of Bernard of Clairvaux*, ed. and trans. Hilary Costello, Cistercian Fathers Series 76 (Collegeville, MN: Liturgical Press, 2015).

[39] Strictly speaking, there is no '*Vita*' for Elisabeth falling under that exact term, but there is an obituary letter written by Ekbert after the visionary's death. See Ekbertus Schoenaugiensis, *De obitu Elisabeth*, in *Die Visionen der hl. Elisabeth und die Schriften der Aebte Ekbert und Emecho von Schönau*, ed. Ferdinand W.E. Roth (Brünn: Raigerner Benediktiner Buchdruckerei. Verlag der "Studien aus dem Benediktiner- und Cistercienser-Orden", 1884), 263–78.

[40] See Godefridus Sancti Disibodi and Theodericus Epternacensis, *Vita sanctae Hildegardis virginis*, ed. Monika Klaes, *CC CM* 126 (Turnhout: Brepols, 1993). The *Vita Hildegardis* and its composition history is assessed anew in this thesis, from p. 177 onward. The text is published by Klaes and is —similarly to Bernard of Clairvaux's *Vita*— the work of multiple authors and redactors. Hildegard and Volmar have likely begun the *Vita* in the earliest phases. Godfrey of Disibodenberg (book I) and Theoderic of Echternach (books II and III) have with certainty contributed to it. The latter is often referred to as ultimate end redactor and main contributor because he has finalized the *Vita* as a whole. There is also a revised version of the *Vita* by Guibert of Gembloux, which forms part of Klaes's appendix in the cited edition on at 91–106. An earlier fragmentary *Vita*, likewise written by Guibert of Gembloux, is part of his letter collection, see Ep. 38 in Guibertus Gemblacensis, *Guiberti Gemblacensis Epistolae quae in codice B.R. BRUX. 5527–5534 inueniuntur*, ed. Albert Derolez, *CC CM* 66–66A (Turnhout: Brepols, 1988–9), 369–79.

[41] Gillian Rosemary Evans, *Bernard of Clairvaux*, Great Medieval Thinkers (New York: Oxford University Press, 2000), 5.

diocesan approval" was the deciding factor of elevation to sainthood.[42] This regulated process for the endowment of sainthood could justify the hiring of professional hagiographers.[43] Hildegard's (auto)hagiography, for instance, the *Vita Hildegardis*,[44] was redacted by a chronicler, Theoderic of Echternach, who had never met the visionary personally, but had enjoyed the skill and education to bring the project to a satisfying conclusion.[45] Fervent appeals were made to Popes Gregory IX (1237), Innocent IV (1243) and John XXII (1317), but the decision was not pushed through officially until 2012.[46] With Bernard of Clairvaux, it was his secretary Geoffrey of Auxerre who was responsible for the compilation and finalization of the abbot's *Vita prima*.[47]

### 1.4.2 Conflict in the Scriptorium

This serves to demonstrate that on behalf of the secretary a writing collaboration meant utmost dedication, and from the author it required a great amount of confidence. It is difficult to fully perceive and comprehend the sensitivities such a collaboration must have brought along. Only once in a while, we are granted a rare glance at the kinds of misunderstandings and jealousies that could arise behind the screens. If the bond of trust was broken, or if a secretary overstepped the line, emotional responses could rise high. One example is the conflict between Hildegard of Bingen and her female collaborator Richardis of Stade. When the latter decided to leave Disibodenberg with the purpose of founding her own abbey, Hildegard's response was a mix of anger and despair, which ultimately resulted in a accusatory letter to Richardis's address.[48] Her complaints would prove to be in vain, and Richardis would leave Hildegard's side nevertheless.

Contrarily, Peter the Venerable's appeals to invert the departure of his secretary Peter of Poitiers (c. 1130–1215) appear to have been more successful. Some time around 1134, secretary Peter left Cluny temporarily to find retreat in solitude. Peter

[42] Sabina Flanagan, *Hildegard of Bingen, 1098–1179: A Visionary Life*, 2nd ed. (London: Routledge, 1998 (1989)), 11–2.

[43] Evans, *Bernard of Clairvaux*, 5.

[44] Discussed in chapter 6 of this thesis, see pp. 177–201.

[45] See the introduction to Godefridus Sancti Disibodi and Theodericus Epternacensis, *Vita Hildegardis*, 77*.

[46] Barbara Newman, "St. Hildegard, Doctor of the Church, and the Fate of Feminist Theology," *Spiritus: A Journal of Christian Spirituality* 13, no. 1 (2013): 36–55.

[47] The last parts of the *Vita* were written by himself, and the first by William of Saint-Thierry and Arnaud of Bonneval. A first appeal to canonization shortly after Bernard's death had been denied. Only in 1174, some twenty years after Bernard's death, Geoffrey would successfully submit a revised *Vita* that granted Bernard his ultimate canonization. An earlier attempt had been made right after Bernard's death, but was rejected by Innocent II. On Bernard's *Vitae* and the various extant versions, see n. 38

[48] "Hildegard, unwilling to accept [the departure of Richardis] , in her doting attachment, made an exhibition of herself in a way she never did (though she had feared to) in her visions," see Peter Dronke, *Women Writers of the Middle Ages: A Critical Study of Texts from Perpetua († 203) to Marguerite Porete († 1310)* (Cambridge, UK: Cambridge University Press, 1984), 154–6; quote taken from 156; also see Sabina Flanagan, "*Spiritualis amicitia* in a Twelfth-Century Convent? Hildegard of Bingen and Richardis of Stade," *Parergon* 29 (1981): 15–21.

the Venerable had, however, been caught in the midst of stressful affairs which direly required Peter's friendship and assistance. Despite the fact that the abbot of Cluny had given his fiat to Peter's departure, he made fierce reproaches in a letter, arguing that his secretary had elected personal needs over his superior. The message came across, and Peter of Poitiers returned, remaining Peter the Venerable's secretary until his death.[49]

In such an atmosphere of privileged positions, where loyalty bordered on sycophancy, envy is never far-removed. In Suger's *Vita*, secretary William of Saint-Denis insinuated that Suger had by the end of his life become encircled by trustees who did not always have the abbot's best interests at heart. Suger's politics of 'favouritism' and exclusivity in selecting those whom he trusted most resulted in conspiracies between distinctive parties of his entourage, which would break out fully after his death.[50] William himself had only later in life been admitted to Suger's inner circle.[51] In a turbulent succession phase after Suger's death († 1151), he would lead an uprising faction of monks against the new abbot and former secretary of Suger, Odo of Deuil, who has attracted suspicions of theft and murder to his person. Ultimately, William's opposition against Suger's other popular secretary led to his banishment from the monastery.[52]

Finally, one more well-known dispute between author and secretary should be mentioned here. Nicholas of Montiéramey, Bernard of Clairvaux's closest secretary, appears to have severely abused Bernard's trust in the final years of Bernard's life, around 1151–2. Bernard addresses the conflict between him and Nicholas in a letter to Pope Eugene III, stating that "Nicholas has left us, because he was not one of us."[53] Apparently, Nicholas had instigated Bernard's disconcertion by sending out letters under his name and seal without approval. It is a curious case,[54] in which very little evidence to ascertain the nature of Nicholas's transgression is available. What is clear, however, is that Bernard's faith had been breached sufficiently to expel his head secretary from Clairvaux.

[49] The letter, Ep. 58, is edited in Petrus Venerabilis, *The Letters of Peter the Venerable*, ed. Giles Constable, Harvard Historical Studies 78 (Cambridge, MA: Harvard University Press, 1967), 2:179–89. For a discussion of Ep. 58, "Ad Petrum Pictavensem suum," and Peter of Poitiers' assistance to the abbot of Cluny, see ibid., 2:331–43.

[50] For William's remark upon jealousy: "Hec ideo dixerim ut sciant emuli, audiant obtrectatores cujus apud reges loci, quante reverentie apud optimates extiterit," see Guillelmus Sancti Dionysii, *Vie de Suger*. The case of Suger's favouritism and his encirclement by those who envied him was most strongly made in Lindy Grant, *Abbot Suger of St-Denis: Church and State in Early Twelfth-Century France*, ed. David Bates, The Medieval World (New York: Addison Wesley Longman, 1998), 205–7.

[51] Hubert Glaser, "Wilhelm von Saint-Denis: Ein Humanist aus der Umgebung des Abtes Suger und die Krise seiner Abtei von 1151 bis 1153," *Historisches Jahrbuch* 85, no. 2 (1965): 257–322.

[52] Discussed more fully on p. 216 in this thesis.

[53] "Nicolaus ille exiit a nobis, quia non erat ex nobis," see Bernardus Claraevallensis, Ep. 298, *SBO* 8:214.

[54] The case of Nicholas's transgression forms the point of departure for a detailed study of his stylistic influence on Bernard's literary output from p. 125 onward.

## 1.5 Female Authorship and Cross-Sex Collaborations

### 1.5.1 The *cura monialium*

A particular kind of writing partnership in this thesis will be that between the sexes.[55] Since medieval culture discouraged women from entertaining learning in the schools and from taking up didactic positions, the few examples of female writers that we know of for the Early and High Middle Ages are almost without exception found in monasteries. That female authorship existed specifically there has significant implications, for female monasticism was highly dependent on the whims of a Church led by men, and the relations between the sexes so happened to have come under constant negotiation. In all aspects of religious life, the presence of women constituted a problem for men. The Benedictine Rule, as Heloise noted in her sixth letter to Abelard, had very little to say for female spirituality.[56] Paul's letters had explicitly silenced women and subordinated them to men.[57]

Whereas within early medieval female convents women religious had experienced a higher degree of liberty as to their practices and rules of conduct,[58] the centuries in the

[55] The bibliography on female writers and their male collaborators is long. A nuanced and important study of the dynamics of power and authority between the female writer and her collaborator is that of John Wayland Coakley, *Women, Men and Spiritual Power: Female Saints and Their Male Collaborators* (New York: Columbia University Press, 2006); Benedict has made an excellent study about cross-gender partnerships as narrative constructions, see Kimberley M. Benedict, *Empowering Collaborations: Writing Partnerships between Religious Women and Scribes in the Middle Ages*, ed. Francis G. Gentry, Studies in Medieval History and Culture 27 (New York: Routledge, 2004).

[56] Heloise's criticisms formulated to Abelard form the departure point in Barbara Newman, "Flaws in the Golden Bowl: Gender and Spiritual Formation in the Twelfth Century," *Traditio* 45 (1989): 111–46. For the passage in the letter collection (both in Latin and English translation), see Petrus Abaelardus and Heloisa Argentoliensis, *The Letter Collection of Peter Abelard and Heloise*, ed. David E. Luscombe, trans. David E. Luscombe and Betty Radice, Oxford Medieval Texts (Oxford, UK: Clarendon Press, 2013), §4, 220–1.

[57] Especially 1 *Timothy* 2.11–5 is often found quoted in this regard.

[58] I am making a vast simplification of a very complex, centuries-long history of female religiosity before Gregorian reform (c. 1050–80), a time which in reality testifies of numerous and various attempts to regulate spiritual life, and which boasts a rich variety of forms of female monasticism. The grand narrative argues that medieval female monasticism before the twelfth century experienced its highest degree of liberty in the formative, experimental phase of the Church during the Merovingian period roughly before ca. 740. See Jane Tibbets Schulenburg, "Women's Monastic Communities, 500–1100: Patterns of Expansion and Decline," *Signs* 14, no. 2 (1989): 261–92; and Suzanne Fonay Wemple, *Women in Frankish Society. Marriage and the Cloister 500 to 900*, The Middle Ages (Philadelphia: University of Pennsylvania Press, 1985); with the Carolingian reform one increasingly observes the active imposition of rules, such as adaptations of the sixth-century Benedictine Rule or the *Institutio sanctimonialium* issued at the Aachen Council of 816, which sought to actively seclude women religious from any kind of political partaking in the structures of the Church. Nevertheless, in light of women's flexible interpretations of these rules in post-Carolingian times, Vanderputten, Thibaut and others have emphasized the danger in unilaterally interpreting Carolingian customaries to reflect the lived reality of women religious at this time. Instead, the decades following the Carolingian reform appear to have seen a greater degree of experimentation than has hitherto been presumed, where women could creatively cope with the prevailing heterogeneity and ambiguous attitudes at the time. See Steven Vanderputten, *Dark Age Nunneries. The Ambiguous Identity of Female Monasticism, 800–1050* (Ithaca and London: Cornell University Press, 2018), 6–8; Jirki Thibaut, "Intermediary Leadership. The Agency of Abbesses in Ottonian Saxony," in *Abbots and Abbesses as a Human Resource in the Ninth- to Twelfth-Century West*, ed. Steven Vanderputten, Vita regularis. Ordnungen und Deutungen religiosen Lebens im Mittelalter 74 (Münster: Lit, 2018), 6.

wake of the eleventh-century Gregorian reform[59] increasingly sought to regulate their ways, and imposed the Benedictine Rule as an attempt to 'standardize' or formalize monastic life for women religious.[60] Amongst other measures, this entailed that male supervisors were installed to keep an eye out for women religious, a phenomenon that has generally been termed the *cura monialium*. Literally this translates to the 'caretaking by monks,' which Golding more generally rephrased as "shorthand for the spiritual oversight in the provision of pastoral care and for the authority exercised over female communities by male clerics."[61] In practice, the *cura* entailed the appointment of male spirituals, provosts and/or confessors, whose function was to answer to the practical and ceremonial needs (the liturgy and the sacraments) of a female or mixed community. One can —and probably should— make a distinction between a female community supervised by a single provost or priest, and 'actual' double houses, in which a male monastery and a female nunnery where housed under a single roof, neighboured each other, or were divided only by a wall.[62] The late eleventh and twelfth century was in a transitory phase when it came to the symbiosis of the sexes. The terms by which men and women's joint cenobitic life was conducted were still loose, and the practices that were stipulated by individual communities still co-existed with the ones dictated by a centralized ecclesiastic institution or orders' distinctive practices. Such was certainly the case until around 1139, the time of the second Lateran council, which more strictly distinguished and demarcated the roles of the sexes in monastic life, and brought to a close "an age of experimentation."[63]

The *cura* forms the backdrop against which the cross-sex collaborations discussed in this thesis need to be understood. Both Elisabeth of Schönau and Hildegard of Bingen (discussed respectively in chapters 5 and 6) wrote their works in a double house, Schönau and Disibodenberg,[64] which had been heavily under the influence of the Hirsau network in the diocese of Speyer. Under the abbacy of William of Hirsau (c. 1030–1091), this abbey situated in the Black Forest had become the centre of reform for a network of German monasteries in the course of the eleventh and twelfth century, and was known to have encouraged the monastic life of women living alongside men.[65] Also Abelard, as founder of the Paraclete, provided Heloise with

[59] Constable, *The Reformation of the Twelfth Century*.

[60] Fiona J. Griffiths, "The Cross and the *Cura monialium*: Robert of Arbrissel, John the Evangelist, and the Pastoral Care of Women in the Age of Reform," *Speculum* 83, no. 2 (2008): 309ff.

[61] Brian Golding, "Bishops and Nuns: Forms of the *cura monialium* in Twelfth- and Thirteenth-Century England," in *Women in the Medieval Monastic World*, ed. Janet Burton and Karen Stöber, Medieval Monastic Studies 1 (Turnhout: Brepols, 2015), 97.

[62] For the different types of male-female symbioses covered by the term 'double house,' see Elm and Parisse, *Doppelklöster*.

[63] Constant J. Mews, "Negotiating the Boundaries of Gender in Religious Life: Robert of Arbrissel and Hersende, Abelard and Heloise," *Viator* 37 (2006): 148.

[64] Only in the beginning of her career. In 1150, Hildegard moved to the Rupertsberg.

[65] See Julie Hotchin, "Female Religious Life and the *Cura Monialium* in Hirsau Monasticism, 1080 to 1150," in *Listen, Daughter. The Speculum Virginum and the Formation of Religious Women in the Middle Ages*, ed. Constant J. Mews,

his *cura monialium*. Abelard wrote sermons, hymns, poems and commentaries at Heloise's request, and procured her with the necessary directives for leading a spiritual life.[66]

### 1.5.2 Repression

Either due to their collaboration with male clerics or for anachronistic or sometimes downright denigrating reasoning, female medieval authors have in the past not seldom been forgotten, neglected, repressed, or flatly denied the authorship over their work. One example could be Hrotsvitha, canoness of Gandersheim (935–c. 1010), whose works were argued to be a humanist fabrication by the nineteenth-century scholar Joseph Aschbach.[67] It turned out, however, that Aschbach had curiously few arguments for supporting this claim, and that he was driven mainly by an implicit conviction that a woman could not have been the author of such great works of Latin literature.[68] In a similar vein, Hildegard of Bingen has been denied authorship over her works in the nineteenth century by Wilhelm Preger.[69] Preger's convictions were able to sustain a considerable amount of doubt in the decades after. Only with the arrival of the seminal study of Marianna Schrader and Adelgundis Führkötter, contending for the *Echtheit* of Hildegard's works, his theses were conclusively disproved.[70]

Perhaps the most fervently discussed case of female authorship is that of Heloise. Despite Heloise's reputation as a brilliant writer in her own day, nearly all of her works have been disputed, mostly in favour of her lover Peter Abelard.[71] Elsewhere,

The New Middle Ages (New York: Palgrave, 2001), 59–83; the conditions of Elisabeth of Schönau's life in the Schönau double house are more adequately described in n. 15 on p. 150. For the Hirsau movement and its influence on Disibodenberg specifically, see Jeroen Deploige, *In nomine femineo indocta: Kennisprofiel en ideologie van Hildegard van Bingen ( 1098-1179 )* (Hilversum: Verloren, 1998), 19; and Constant J. Mews, "Hildegard of Bingen and the Hirsau Reform in Germany 1080-1180," in *A Companion to Hildegard of Bingen*, ed. Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, Brill's Companions to the Christian Tradition 45 (Leiden: Brill, 2014), 57–83.

[66] Fiona J. Griffiths, "'Men's Duty to Provide for Women's Needs': Abelard, Heloise, and their Negotiation of the *cura monialium*," *Journal of Medieval History* 30, no. 1 (2004): 1–24; and Fiona J. Griffiths, "Monks and Nuns at Rupertsberg: Guibert of Gembloux and Hildegard of Bingen," in *Partners in Spirit: Women, Men, and Religious Life in Germany, 1100-1500*, ed. Fiona J. Griffiths and Julie Hotchin, Medieval Women: Texts and Contexts 24 (Turnhout: Brepols, 2014), 149. On the works Abelard wrote at the request of Heloise, see p. 254 further down this thesis.

[67] See Joseph Aschbach, *Roswitha und Conrad Celtes* (Vienna: K.K. Hof- und Staatsdruckerei, 1867), According to Aschbach, Celtis's aim in forging the works of Hrotsvitha would have been to glorify the literary accomplishments of the German Middle Ages, to promote a German muse, and to elevate its achievements to be on par with that of Italian literature.

[68] "Dazu kommt, dass der Geist, der diese Werke durchweht, durchgehends ein männlicher und sehr gebildeter ist," see Aschbach, *Roswitha und Conrad Celtes*, 21; criticism on Aschbach's flawed critical attitude to the sources was, however, soon to follow in that same century, see William Henry Hudson, "Hrotsvitha of Gandersheim," *The English Historical Review* 3, no. 11 (1888): 431–57.

[69] Discussed more fully on p. 177 in this thesis. See Wilhelm Preger, *Geschichte der deutschen Mystik im Mittelalter*, vol. 1 (Leipzig: Dörffling und Franke, 1874), 29–36.

[70] Marianna Schrader and Adelgundis Führkötter, *Die Echtheit des Schrifttums der Heiligen Hildegard von Bingen* (Cologne / Graz: Böhlau Verlag, 1956).

[71] For a full discussion and listing of Heloise's works, see pp. 249ff. A sharp and witty publication summarizing some of the most ridiculous scholarly refusals in granting Heloise authorship over her texts, can be found in Barbara Newman, "Authority, Authenticity, and the Repression of Heloise," *Journal of Medieval and Renaissance Studies* 22, no. 2 (1992):

we find that scholars only allow women writers to have a say in their own texts when that is stereotype-confirming. Roth, when discussing Elisabeth of Schönau's visions, argued that if the latter's presence was anywhere to be discerned, it was in those passages most 'hysterical,' as male writers would not be induced to describe such ecstatic, dreamlike and fanciful images.[72]

### 1.5.3 Recovery

The past decades have witnessed a counter-movement, in which the anachronistic and sometimes blatantly misogynistic motivations for denying women their own authorship have been gradually debunked, and the mission is still on course.[73] Ever since, a prominent place in the debate on medieval authorship is reserved for female authors' agency and creativity, whose works have in the past often been ascribed to male writers or male secretaries. The revaluation of the female role in medieval literary production and their involvement in the literary canon has borne beyond the obvious rectifications for a Western-European history of female writing. It has in the meantime become part of a more comprehensive and expanded understanding of medieval authorship in general. So, for instance, abbess Herrad of Hohenbourg's *Hortus deliciarum* (c. 1130–1195), a compilation of texts from Latin authors such as Augustine, Isidore of Seville, Peter Lombard or Rupert of Deutz, is now recognized to be an authorial creation, in that she glossed (in German) and decorated the material she had collected with her fellow sisters. In reading the *Hortus deliciarum* one sees that the compilations were a selective and meticulous process, worthy of being regarded a form of authorship in its own right with a rationale and didactic programme. Parts were extracted, others retained, all of which required reading and understanding, and reinterpretation through reordering. This, also, is medieval authorship.[74]

Medieval female writers have been shown to have been most creative in playing out their subordinate position to their advantage. Not coincidentally, two of the three female authors discussed in this thesis were visionary writers, Heloise being the exception to Hildegard of Bingen and Elisabeth of Schönau.[75] Prophecy has been considered a specifically 'female' genre, providing "a direct and privileged relation with

121–57.

[72] "Vor allem das innere Seelenleben, dass aus ihnen spricht, das Extase, die melancholischen Schmerzen, hysterischen Verzückungen, wie solches alles nur dem Weibe eigen ist [...] Dieses alles schützt die Visionen, deren einfache Angaben zu wahrheitsgetreu sind, um Mächwerk eines Mannes zu sein, vor allen Angriffen der Unächtheit," see Ferdinand W.E. Roth, ed., *Die Visionen der hl. Elisabeth und die Schriften der Aebte Ekbert und Emecho von Schönau* (Brünn: Raigerner Benediktiner Buchdruckerei. Verlag der Studien aus dem Benediktiner- und Cistercienser-Orden, 1884), xcv.

[73] Anne L. Clark, "Repression or Collaboration? The Case of Elisabeth and Ekbert of Schönau," in *Christendom and its Discontents. Exclusion, Persecution, and Rebellion, 1000-1500*, ed. Scott L. Waugh and Peter D. Diehl (Cambridge, UK: Cambridge University Press, 1996), see especially 151–2.

[74] Herrad Hohenburgensis, *Hortus deliciarium*, ed. Rosalie Green et al., Studies of the Warburg Institute 36 (Leiden: Brill, 1979).

[75] A basic work of study on visionary literature remains Peter Dinzelbacher, *Vision und Visionsliterature im Mittelalter*, Monographien zur Geschichte des Mittelalters 23 (Stuttgart: Hiersemann, 1981).

God *ex gratia*, which frequently expressed itself in paramystical experiences usually visible to their audience such as prophecy, glossolalia, visions, apparitions, and ecstasies."[76] In other words, female authorship drew its authority from its revelatory quality, boasting a spiritual, charismatic connection to God that was more informal and based on experience than that of male clerics, whose authority was mainly situated in office and learning.[77] Hildegard of Bingen might serve as one of the earliest examples of appropriating this kind of authoritative position for herself, an example which would soon be followed by a long list of well-known holy women such as Hadewich of Brabant (before 1250), Mechtild of Magdeburg (c. 1207–c. 1282/94), Angela of Foligno (1248–1309), Marguerite Porete († 1310), Birgitta of Sweden (1303–73), Julian of Norwich (c. 1342–c. 1416), Catherine of Siena (1347–80), Margery Kempe (c. 1373–c. 1440), etc.

Both Hildegard and Elisabeth are considered to have strategically exaggerated the deficiencies and self-devaluations to which their sex had condemned them. In the quote given earlier in this chapter on p. 34, it becomes clear how Hildegard exploited her unique celestial privileges to reinforce the demand that not her collaborators nor anyone else adjusted the contents of her revelations. The word of God was not to be meddled with, and by invoking this ultimate authority Hildegard and Elisabeth were able to place their male collaborators before a stalemate. At work here is a delicate balancing act between self-devaluation and self-authorization, between proclaiming insufficiency and simultaneously asserting divinely inspired charisma.

## 1.6 *Auctoritates*

### 1.6.1 The Bible and the Church Fathers

The author not merely collaborates with physical assistants or partners, but also with a past. Traditionally, there is no medieval authorship imaginable without reference to Scripture and patristic heritage. The importance for a twelfth-century author to place him- or herself in line with a past of culture and learning, manifests itself in extensive borrowings and phrasings from authoritative sources. These authorities —

[76] *Ex gratia* is consequently often found opposed to male authority that was institutionalized and therefore *ex officio*. See Veerle Fraeters and Imke de Gier, "Introduction: Shaping Female Spiritual Authority in Europe from the High Middle Ages to the Early Modern Period," in *Mulieres religiosae. Shaping Female Spiritual Authority in the Medieval and Early Modern Periods*, ed. Veerle Fraeters and Imke de Gier, Europa Sacra 12 (Turnhout: Brepols, 2014), 3. All generalizations must come with their nuances. A study by Deploige explored the case of Simon of Aulne († 1229), a male clairvoyant. See Jeroen Deploige, "How Gendered Was Clairvoyance in the Thirteenth Century? The Case of Simon of Aulne," in *Speaking to the Eye: Sight and Insight through Text and Image (1150-1650)*, ed. Thérèse de Hemptinne, Veerle Fraeters, and María Eugenia Góngora, MISCS 2 (Turnhout: Brepols, 2013), 95–126.

[77] In this department, especially the recent work of Coakley deserves mentioning, see Coakley, *Women, Men and Spiritual Power*; and John Wayland Coakley, "Women's Textual Authority and the Collaboration of Clerics," in *Medieval Holy Women in the Christian Tradition (c. 1100–c.1500)*, ed. Alastair Minnis and Rosalynn Voaden, Brepols Essays in European Culture 1 (Turnhout: Brepols, 2009), 83–104.

especially in Latin— largely prescribe and regulate the norms and traditions connected to authorship.

God functions as the ultimate *auctor*, and his *verbum* as the source from which all human knowledge emanates. This idea of the human author being fully dependent upon God is a commonly found trope in medieval texts. So, for example, Anselm of Canterbury stated: "I am certain that if I say anything that is indubitably contradicted in Sacred Scripture, it is a falsitude."[78] When speaking of interpreting her visions, Hildegard of Bingen describes herself as a feather in the wind, deprived of all independent power and sustained by the gust of God only.[79] It is not Hildegard who interprets (and who writes), but it is God, and she is but the vessel. In a similar vein, Bernard of Clairvaux's texts strike us as receptacles of divine inspiration, where clear reminiscences of Biblical writing are cleverly composed into a new whole. The Cistercian abbot does not simply 'cite' the Bible, nor Augustine and Ambrose for that matter.[80] He is immersed and indebted to the extent that he speaks and writes 'biblically.' The end result is nevertheless always very distinctively 'Bernardian.'

This has important implications for how the text was seen in relationship to the physical author. It also raises significant questions concerning stylistic affinities standing central in this thesis. To what extent does this persistent presence of a subtext shroud the individuality of the author in terms of personal style? Or, can we turn this idea around: to what extent does selection —the preference of certain fragments and their encapsulation within a new, self-made script— form a component of an author's personal creativity? G. Udny Yule, a forefather of computational stylistics with an interest for the Middle Ages (see below on p. 64), spoke of the troubles such stylistic indebtedness brought along for the 'formal' recognition of text by the fifteenth-century author Thomas à Kempis (c. 1380–1471):

> Thomas à Kempis, for example, quotes the words of scripture so freely that if one cut out scriptural quotations one would eliminate a considerable proportion of his work. He has made scripture his own, and what he has written must stand as his.[81]

## 1.6.2 Compliance to Authority

Text that appealed to authority, and incorporated authoritative cultural artefacts, universally transcended particular or individual forms of authority. The collective decided

[78] My translation. "Certus enim sum, si quid dico quod Sacrae Scripturae sine dubio contradicat, quia falsum est," in Anselmus Cantuariensis, *Cur deus homo*, in *S. Anselmi Cantuariensis archiepiscopi opera omnia*, ed. Franciscus Salesius Schmitt, vol. 2 (Apud Thomam Nelson et Filios, 1946), 1.18.8, 82.

[79] "[...] Quatinus uelut penna, que omni grauedine uirium caret et que per uentum uolat, ab ipso sustinear," in Godefridus Sancti Disibodi and Theodericus Epternacensis, *Vita Hildegardis*, 1.8, ll. 16–7, 14.

[80] Christine Mohrmann, "Observations sur la langue et le style de saint Bernard," in *SBO 2*, ed. Jean Leclercq, Henri M. Rochais, and Charles Holwell Talbot (Rome: Editiones Cistercienses, 1958), xiii.

[81] George Udny Yule, "On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship," *Biometrika* 30, nos. 3/4 (1939): 366.

what was regarded authoritative and therefore worthy of attaining and exercising *auctoritas*.[82]

Ziolkowski stressed the difficulty of that term, and emphasized that *auctoritas* entails more than reference to authoritative individuals —or *auctores*—, but extends towards text and language. The concept was originally Roman, and it never had a univocal meaning in the Middle Ages. Many connotations persist: "impressiveness of style" or "normative literary usage," "people of authority" and "extracts from texts that confer authority."[83] For Minnis "the writings of an *auctor* contained, or possessed, *auctoritas* in the abstract sense of the term, with its strong connotations of veracity and sagacity. In the specific sense, an *auctoritas* was a quotation or an extract from the work of an *auctor*."[84]

One will often find *auctoritas* discussed in conjunction with a well-known twelfth-century metaphor attributed to Bernard of Chartres. The metaphor is cited in John of Salisbury's († 1180) *Metalogicon* and compares authors to "dwarfs standing on the shoulders of giants."[85] Text was always considered a continuation or elaboration of what had already been written. Re-use of text did not reflect a lack of originality or inspiration. It was a *sine qua non*, a commitment to and acknowledgement of contemporary intellectual readership. The 'dwarfs'-metaphor of Bernard of Chartres —although its meaning is much disputed—,[86] generally represents the humility and indebtedness of the 'moderns' —the dwarfs— in face of the 'ancients' —the giants.[87] And yet this is not a simple, polite recognition of the tradition. Along the way, Bernard of Chartres's metaphor also implies an idea of progressive optimism. The dwarfs can, after all, see wider and further than their predecessors. In this context of germinating humanist ideals and early rationalization of thought, the notions of *auctoritas* and authority were under negotiation.

[82] See Masters, "Distribution, Destruction and Dislocation," 275–6.

[83] Jan M. Ziolkowski, "Cultures of Authority in the Long Twelfth Century," *The Journal of English and Germanic Philology* 108, no. 4 (2009): 426.

[84] Minnis, *Medieval Theory of Authorship*, 10–2.

[85] "Dicebat Bernardus Carnotensis nos esse quasi nanos gigantum umeris insidentes, ut possimus plura eis et remotiora uidere, non utique proprii uisus acumine, aut eminentia corporis, sed quia in altum subuehimur et extollimur magnitudine gigantea," see Joannes Saresberiensis, *Metalogicon*, ed. John Barrie Hall and K. S. B. Keats-Rohan, *CC CM* 98 (Turnhout: Brepols, 1991), 3.4.46.

[86] Edouard Jeauneau, "'Nani gigantum humeris insidentes'. Essai d'interprétation de Bernard de Chartres," *Vivarium* 1 (1967): 79–99.

[87] The word 'modernus' would increasingly surface in the twelfth century. See Elisabeth Gössmann, "'Antiqui' und 'moderni' im 12. Jahrhundert," in *Antiqui und moderni. Traditionsbewußtsein und Fortschrittsbewußtein im späten Mittelalter*, ed. Albert Zimmermann, Miscellanea mediaevalia 9 (Berlin: De Gruyter, 1973), 41; and Richard William Southern, *Scholastic Humanism and the Unification of Europe*, vol. 1. Foundations (Cambridge, MA: Blackwell, 1997 (1995)), 185.

### 1.6.3 Questioning of Authority

Alan of Lille (c. 1120–1202), in defending catholic faith in his *Contra haereticos*,[88] argued that "authority has a waxen nose, that can be bent in different ways" and was to be "fortified by reasons."[89] The implication is that *auctoritates* were interpretable and required the support of human *ratio* in order not to become abused by those whom Alan perceived to be heretics. Especially in the emerging schools in and around Paris, one finds clear examples of how the 'old' authorities were put to the test. Doctrine was in dire need of organization and systematic clarification. By help of the ongoing literacy and the secularization of school systems, philological techniques arose in which a place was reserved for human reasoning to distinguish the authoritative from the apocryphal.[90] So the Benedictine Rule was scrutinized from up closer, which came with a realization that the contemporary practice of the Church did not live up to its true tenets.[91] Theological issues were investigated and disputed, as is attested by the numerous treatises on the nature of the Trinity. The controversies that could arise in such theological issues —such as, indeed, the Trinity— is best epitomized by the accusation of heresy directed at Peter Abelard's *Theologia 'summi boni'* at Soissons in 1121, and the confrontation with Bernard of Clairvaux at the Council of Sens in 1141.[92] Abelard also listed what he experienced to be contradictions in the Church Fathers' writings in his monumental *Sic et non* (see also p. 82). The criticism to which authoritative texts were subjected, however, was not limited to the schools. In the monasteries one likewise finds an increased sensitivity to categorize and distinguish the authoritative from the false. For example, when diggings in the North of Cologne brought to the surface the bodies of Saint Ursula and her virgin companionship, Elisabeth of Schönau expressed skepticism over who authored the corpses' name plates.[93] In one of her vision books she subjects an Ursuline virgin appearing before her to a severe interrogation: "My lady, what does it mean that bodies of bishops were also found buried in the place of your martyrdom? And should we believe the titles inscribed on the stones found there? And who wrote them?"[94]

[88] *PL* 210:305–428.

[89] "Sed quia auctoritas cereum habet nasum, id est in diversum potest flecti sensum, rationibus roborandum est," taken from *PL* 210:333. Translation taken from Peter Dronke, ed., *A History of Twelfth-Century Western Philosophy* (Cambridge, UK: Cambridge University Press, 1988), 7.

[90] Curtius, *European Literature and the Latin Middle Ages*, 129.

[91] Morris, *The Discovery of the Individual, 1050-1200*, 57–8.

[92] The clash between Abelard and Bernard has often been seen as a culminating point of a larger conflict between two movements, respectively the 'scholastic' and the 'monastic.' In the meantime, this implication has been contested and exposed for a simplification. See Constant J. Mews, "The Council of Sens (1141): Abelard, Bernard, and the Fear of Social Upheaval," *Speculum* 77, no. 2 (2002): 342–82; Wim Verbaal, "The Council of Sens Reconsidered: Masters, Monks, or Judges?," *Church History* 74, no. 3 (2005): 460–493.

[93] The role Elisabeth played in the authentication of the unearthed Ursuline relics are discussed more fully in this thesis from p. 170 onward.

[94] "Domina mea, quid hoc sibi vult, quod in loco martirii vestri etiam corpora, episcoporum sepulta inveniuntur? Et nunquid credendum est superscriptionibus titulorum, qui in quibusdam lapidibus illic reperiuntur? Et quis fuit scriptor eorum?"The text is taken from her *Liber revelationum* in Elisabeth Schoenaugiensis, *Visionen*, §5, 125. Translation taken

The example of Elisabeth of Schönau is just one out of many examples testifying that the critical attitudes which were destabilizing the *auctoritates* are not only found in North-Western France, nor only in the schools, nor only by men. "Scholastic humanism," then, as it was famously termed by Southern,[95] was a general movement. Its "terms, doctrines and motifs" would circulate and exert influence outside of the schools.[96] As this attitude came in sway, a paradox emerged. On the one hand, written records gained in authority. Yet on the other hand their potential to deceive and manipulate —in a language that was no one's mother tongue any longer—[97] became clearer than ever before.

### 1.6.4 Rediscovery of the Classics

This rearrangement and rethinking of the established medieval Latin canon's *auctoritates* announced a momentum for change. A reacquaintance with familiar pagan authors and a first acquaintance with the vernacular literatures threw the literary canon further out of balance.[98] The earlier cloister schools such as those in Carolingian times had a very pragmatic relationship with the ancients. It was based primarily on grammar and rhetoric and was restricted to didactic purposes. Texts by writers of the stature of Vergil and Cicero, for instance, would have served as technical examples in a liberal arts curriculum, by whose example students were expected to model their Latin style. With the emergence of dialectics in the eleventh- and twelfth centuries, and particularly the rediscovery of Ovid's love poetry, this practical attitude towards the classics shifted to intellectual engagement. Both Ziolkowski and Minnis note that *auctoritas* was not only conferred by an *auctor* or by the content of that *auctor*'s texts. *Auctoritas* was also related to form and style. As Latin became a language of the schools and had to defend its status with regards to the rising vernacular languages, higher demands were set to how Latin should be written. The curricula in the schools took special care to tend to the demand. Latin increasingly became the language of bureaucrats and the higher educated. Obedience to these norms was deemed important. It enhanced the authenticity and reliability of a piece of writing. The importance of this formal *auctoritas*, so to speak, can be witnessed in the broad influence of *artes dictaminis* and *artes praedicandi* (discussed earlier) which regulated how Latin — especially 'authentic' Latin— was written and spoken.

This deepened engagement and model set by classical authors came with a problem,

from Elisabeth Schoenaugiensis, *Complete Works*, 216.

[95] Southern, *Scholastic Humanism*.

[96] Stover makes this argument for Hildegard of Bingen in Justin A. Stover, "Hildegard, the Schools, and Their Critics," in *A Companion to Hildegard of Bingen*, ed. Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, Brill's Companions to the Christian Tradition 45 (Leiden: Brill, 2014), 111.

[97] James J. Murphy, "The Teaching of Latin as a Second Language in the 12th Century," *Historiographia Linguistica* 7, no. 1 (1980): 159–75.

[98] Curtius, *European Literature and the Latin Middle Ages*, 459.

in the sense that it placed the *auctoritas* of biblical and/or patristic Latinate models under pressure. The classical authors set 'higher' norms that stood in contrast with the 'humble' Latin of the Bible. This tension between two styles, the first 'humble' and the other 'sublime,' forms the spine of Erich Auerbach's *Mimesis*, in which the confrontation of the two is what lies at the basis of European literary history.[99] In many senses, the twelfth-century's identification as 'experimental' can be understood from a very similar clash. 'Experimental' can indicate transgression and a renunciation of former *auctoritates*. Simultaneously, Latin's status of being the language of legitimacy, divinity and truth was endangered. The Goliards, with their playful and satirical interaction with the classics, may be accounted as an early example of this development. In the late twelfth and early thirteenth century, poetry surfaces that explicitly wishes to identify as new, as *modernus*. The canon of *auctoritates* is reformed and contains contemporaries, as becomes clear from such works as Matthew of Vendôme's *Ars Versificatoria* (written some time before 1175), John of Hauville's *Architrenius* (c. 1184), Gervase of Melkley's *Ars poetica* (c. 1185–?), Geoffrey of Vinsauf's *Poetria Nova* (c. 1200), and John of Garland's *Parisiana poetria de arte prosaica, metrica, et rithmica* (1220–35).[100]

These writers are situated in a period that lies safely ahead of the period which is our current focus, and even in their own time they can be regarded something of an exception. The question is, however, which precedents of authorship cleared the stage for them. What kind of authorship could have given rise to these far more unambiguous examples of wayward writing?

# 1.7 Plagiarism, Forgery and Fiction

## 1.7.1 Imitation and Impersonation

A recurring question surfacing throughout this thesis will be to what extent authors tolerated imitation, and at which point impersonation was experienced as infringement. In a textual culture that condoned impersonation, text re-use, anthologization and compilation to such a large degree as that of the Middle Ages, this is not always an easy question. Clearly the terms were defined more loosely than in the centuries postdating the arrival of the printing press in the fifteenth century, traditionally regarded the conception date of notions such as intellectual property and copyright.[101] Honorius Augustodunensis even appears to have deliberately given up the only form of

[99] Erich Auerbach, *Mimesis: The Representation of Reality in Western Literature*, 50th ed., trans. Willard R. Trask (Princeton: Princeton University Press, 2003 (1946)).

[100] Janet Martin, "Classicism and Style in Latin Literature," in *Renaissance and Renewal in the Twelfth Century*, ed. Robert L. Benson, Giles Constable, and Carol Dana Lanham (Toronto: University of Toronto Press, 1991 (1982)), 539–40.

[101] Ziolkowski, "Cultures of Authority," 423.

copyright available to medieval authors: his name.[102] And yet, a considerable number of counterexamples suffice to argue that titled authority, unviolated text and personal style were nevertheless cultivated in the twelfth century.[103] One may presume that, aside from such factors as language and genre, much depended upon the authority at hand, Bernard of Clairvaux would have exerted a very different kind of authority toward his secretaries than Elisabeth of Schönau when his texts were to be redacted and revised. In other words, as we will come to discuss throughout this thesis, writing partnerships were monitored by a set of literary-social norms and values which demarcated the respective spheres of influence, and distinguished intrusions that were considered acceptable from those that were considered transgressive.

### 1.7.2 The (Un-?)intentionality of Deception

Scholars have long been troubled by the aspect of 'plagiarism' and 'forgery' in medieval literature, both of which are blunt and inadequate terms. In fact, in order to discuss what either of these mean in particular contexts, one is in need of a universally valid definition, which Umberto Eco has demonstrated to be a matter far from straightforward.[104] In their most basal definitions, plagiarism means to take some one else's ideas or words and to pass them off as one's own, whereas forgery is the opposite: to take one's own words and pass them off as someone else's.[105] A simple but pertinent question which one might ask is: "When does a letter drafted by a secretary or painting finished by an assistant become a forgery?"[106]

As of late many scholars have emphasized that terms such as plagiarism and forgery have misguiding connotations with deception, criminality and appropriation, which are unbefitting for a medieval context. Dronke, in discussing the letter collection of

[102] In the preface to the *Elucidarium*, *PL* 172:1110. "Nomen autem meum ideo volui silentio contegi, ne invidia tabescens suis juberet utile opus contemnendo negligi: quod tamen lector postulet ut in coelo conscribatur nec aliquando de libro viventium deleatur."The text translates to "I decided to conceal my name, for fear that destructive envy might bid its devotees scorn and neglect a useful work. May the reader, however, pray that it be recorded in heaven, and never be expunged from the book of life."Translation taken from Eva Matthews Sanford, "Honorius, Presbyter and Scholasticus," *Speculum* 23, no. 3 (1948): 399. It should be noted that it was fear for persecution and not modesty that likely drove Honorius to anonymizing his works. On Honorius' disputed identity and provenance, see further down this thesis on p. 105, n. 28.

[103] Griffin makes the case that the chronology of authorship from 'anonymous' in premodern times to a professional author defined by name after the arrival of the printing press is a simplification in both directions. See Robert J. Griffin, "Anonymity and Authorship," *New Literary History* 30, no. 4 (1999): 877–95; Seth Lerer challenged the notion that anthologization and compilation are as strictly medieval as often presumed in Seth Lerer, "Medieval English Literature and the Idea of the Anthology," *PMLA* 118, no. 5 (2003): see especially 1255.

[104] Eventually the discussion leads back to defining the universal meaning of truth and falsehood. Finding such a definition is more than just philosophical meandering, and actually forms an essential point. Most of the confusion and disagreement about medieval plagiarism and forgery might in reality have its roots in a semantic problem, a lack of common ground upon which the topic is discussed. See Umberto Eco, "The Original and the Copy," in *Understanding Origins*, ed. Francisco J. Varela and Jean-Pierre Dupuy, Boston Studies in the Philosophy of Science 130 (Dordrecht: Kluwer Academic Publishers, 1992).

[105] See Giles Constable, "Forgery and Plagiarism in the Middle Ages," *Archiv für Diplomatik* 29 (1983): 3.

[106] Ibid.

the twelfth-century secretary Peter of Blois (c. 1130–c. 1211), which is teeming with what Southern called "terrible plagiarism,"[107] noted that "when applied to twelfth-century writings, this expression is at best an imprecise umbrella-term, at worst an anachronism."[108] Jaeger spoke in similar terms regarding Nicholas of Montiéramey's 'theft' of some of Bernard of Clairvaux's sermons.[109] With diplomats such as Peter of Blois and Nicholas of Montiéramey, the will to imitate and exercise a style in the formulations of the *artes dictaminis* and other forms of school rhetoric (as discussed on p. 27 earlier) are never far off.

When it comes to forgeries, we know that John of Salisbury might have indulged in composing them, but only under the assumption that his hoax would have been recognized by the educated audience he envisioned.[110] Here forgery becomes a literary game, and expects its readership to comply to a certain cultural-literary horizon. Readers who fail to react adequately to John's provocations and accuse him of forgery might simply be –in Constable's terms– "deceived by their own ignorance."[111] Scholars have also justified medieval forgeries by emphasizing the forgers' proper conviction that they were serving a greater good, or that their actions were even divinely endorsed. In discussing secretary Volmar's forgery of a papal letter heading Hildegard's *Epistolarium*,[112] Newman stated that "to Hildegard's contemporaries, including Bernard of Clairvaux and his secretaries, the production of scrupulously revised and idealized letter-collections was not falsification, but a pious work to the glory of God and his saints."[113]

Jean Mabillon (1632–1707) devised guidelines for how to detect forgeries in his *De re diplomatica* (1681).[114] Nineteenth-century philology morally condemned medieval

[107] Peter of Blois's letter collection is full of extensive and often verbatim borrowings taken from elsewhere. See Richard William Southern, *Medieval Humanism and Other Studies* (Oxford: Blackwell, 1970), 107.

[108] See Peter Dronke, "Peter of Blois and Poetry at the Court of Henry II," *Mediaeval Studies* 38 (1976): 199, n. 51.

[109] "Here then is a case in which a skilled student of the *ars dictaminis* with alleged inclinations to forgery imitated a near-contemporary model, and we can assume that there would have been little difference between the 'honest' and dishonest imitation of Bernard's style," in C. Stephen Jaeger, "The Prologue to the *Historia Calamitatum* and the 'Authenticity Question'," *Euphorion* 74, no. 1 (1980): 13. The case of Bernard and Nicholas is fully discussed in chapter 4, from p. 125 onward.

[110] Harold Love, *Attributing Authorship. An Introduction.* (New York: Cambridge University Press, 2002), 179. I am referring to the *Institutio Traiani* supposedly written by Plutarch to the emperor Trajan, a *speculum* for an emperor, which John of Salisbury incorporated in books 5 and 6 of his *Polycraticus*. The forgery is discussed more fully on p. 232.

[111] Giles Constable, "Forged Letters in the Middle Ages," in *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, vol. 5, Monumenta Germaniae Historica Schriften 33, primary (Hannover: Hahnsche Buchhandlung, 1988), 15.

[112] John Van Engen, "Letters and the Public Persona of Hildegard," in *Hildegard von Bingen in ihrem historischen Umfeld. Internationaler wissenschaftlicher Kongreß zum 900jährigen Jubiläum, 13. – 19. September 1998, Bingen am Rhein*, ed. Alfred Haverkamp and Alexander Reverchon (Mainz: Verlag Philipp von Zabern, 2000), 380. The spurious letter is edited by Van Acker in Hildegardis Bingensis, *Epistolarium Pars Tertia CCLI–CCCXC*, 173.

[113] Barbara Newman, "Three-Part Invention: The *Vita S. Hildegardis* and Mystical Hagiography," in *Hildegard of Bingen: The Context of Her Thought and Art*, ed. Charles S.F. Burnett and Peter Dronke, Warburg Institute Colloquia 4 (London: Warburg Institute, 1998), 195–6.

[114] Mabillon is discussed more fully in the next chapter, see p. 85. Christopher N. L. Brooke, "Approaches to Medieval Forgery," *Journal of the Society of Archivists* 3, no. 8 (1968): 378.

literature for its lack of originality and deceptive aims, and concluded that the Dark Ages had brought forth more criminals than any other. Frederick Tout's paper on "Mediæval Forgers and Forgeries" begins in 1920 as such:

> The criminal is with us throughout the ages. He is with us still, though in much reduced numbers, for the farther we go back in history the more criminals we find. In the Middle Ages the criminal class mustered strongly. Not only were mediæval criminals more numerous than their modern counterparts, but by reason of their numbers and importance they excited much more general sympathy than they do nowadays, and were as a rule dealt with by society in a more lenient matter. This was true both of crimes of violence and crimes of deceit. In these two typical classes of misdeeds homicides and forgeries easily took the first places.[115]

Few would subscribe to Tout's opinions today, yet simultaneously, not all scholars have been won for a completely 'relativist' interpretation of —for lack of better terminology— medieval plagiarism and forgery. A scholar such as Brown disagreed with the notion that judging a forger's actions as criminal is motivated by anachronistic criteria.[116]

One may indeed think of those medieval forgeries that strictly appear to have had personal gain at heart. Suger of Saint-Denis, for instance, has been held responsible for producing a dubious charter, allegedly issued by Louis I († 840), by which he was able to expel Heloise and her sisters from Argenteuil and gain possession of their monastery.[117] Clearly Suger drew his inspiration from the rich *chartrier* of Saint-Denis, containing numerous Carolingian diplomas. In such cases, excusing Suger by referring to the 'shared' culture of the Middle Ages, where text was freely appropriated and reintegrated, indeed starts to resemble a skilful dodge in protecting oneself from the embarrassment of catching red-handedly an admirable and important twelfth-century writer in the act of deception. As Michael Clanchy remarked, "forgery [...] was condoned by the greatest scholars, prelates and administrators,"[118] and Constable noted that especially in the twelfth century it increasingly became a source for concern.[119] Simultaneously medieval writers stuck to their conviction that truth was to be fostered and pursued, and lambasted the forging of relics.[120]

[115] Thomas Frederick Tout, "Mediaeval Forgers and Forgeries," in *The Collected Papers of Thomas Frederick Tout with a Memoir and Bibliography*, vol. 3, Lectures, Reprint (Manchester, UK: Manchester University Press, 1934 (1920)).

[116] "What evidence, then, can be cited to show that people in the Middle Ages denied the reprehensibility of forgery or were convinced by persuasive justifications that unimpeachable motivations excused the act of forgery? None that I have seen." See Elizabeth A.R. Brown, "*Falsitas pia sive reprehensibilis*. Medieval Forgers and their Intentions," in *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, vol. 1, Monumenta Germaniae Historica Schriften 33 (Hannover: Hahnsche Buchhandlung, 1988), 106. Brown's arguments are discussed more extensively in chapter 7 on p. 231.

[117] Discussed more fully on p. 214. See Thomas G. Waldman, "Abbot Suger and the Nuns of Argenteuil," *Traditio* 41 (1985): 239–72.

[118] Clanchy, *From Memory to Written Record, England 1066–1307*, 319.

[119] "There seems to have been a rising tide of concern over forgery in the twelfth century." See Constable, "Forgery and Plagiarism," 18.

[120] Guibert of Nogent's polemic *De sanctis et eorum pigneribus* may serve as an example. Guibert criticizes the nearby

### 1.7.3 Fictionality

The lines between imitation, impersonation and taking on a persona are flimsy. Perhaps not coincidentally, the twelfth century witnesses a convergence of the rise of forgery (as well as its counter-reactions) with the rise of literary fiction.[121] This has always placed medieval scholars in the difficult position of having to get a grasp on two important developments simultaneously, which, in a way uniquely medieval, cross each others' domains and interweave fact and myth. Since one would be fully justified in deeming it dangerous to presume that forgeries are always fictions or vice versa, a possible solution for distinguishing them might lie in knowing the author's intention. But this is no easy task. Medieval authors are not exactly known to tip their hand easily. Moreover, along the way, our modern conceptions of what distinguishes truth from falsehood and fact from fiction may betray us in our attempt to expose the author's intention.

In retrospect of what this chapter has thusfar covered, one may argue that many techniques of 'distancing' —which Jean Leclercq would refer to as a "screen of rhetoric"— [122] are simultaneously at work in twelfth-century Latin literature. One may consider the following: the twelfth-century author is an imitator, with (multiple?) secretaries as his or her impersonator(s), writing in Latin —an artificial language neither parties would have spoken as a mother tongue—, collectively lending a voice to a persona within a text which represents neither author nor secretary one-to-one, but which plays part in a purely auto-referential textual construction "no longer necessarily connected to its historical, textual reality."[123] Well-known twelfth-century 'autobiographical' accounts, such as Peter Abelard's *Historia Calamitatum* and Guibert of Nogent's *De vita sua*, truly are anything but autobiographical, or at least not according to its current-day definition. These are not spontaneous, reliable, personal, direct and historical ego-documents. Despite all sense of inwardness and individuality, medieval self-presentations are stylized, designed, by the literal meaning of *fingere* —"formed."[124] Behind their expression of self lies the constant presence of mimesis.[125] The problems

located monastery of Saint-Médard which proclaimed to possess Christ's baby tooth. See Guibertus de Novigento, *De sanctis et eorum pigneribus*, in *Quo ordine sermo fieri debeat; De bucella Iudae data et de veritate dominici corporis; De sanctis et eorum pigneribus*, ed. and trans. Robert B.C. Huygens, *CC CM* 127 (Turnhout: Brepols, 1993), 79–175.

[121] Verbaal, "How the West was Won by Fiction."

[122] Jean Leclercq, "Modern Psychology and the Interpretation of Medieval Texts," *Speculum* 48, no. 3 (1973): 476.

[123] See Verbaal, "How the West was Won by Fiction," 194; and Wim Verbaal, "Trapping the Future: Abelard's Multi-Layered Image-Building," in *Rethinking Abelard: A Collection of Critical Essays*, ed. Babette S. Hellemans, Brill's Studies in Intellectual History 229 (Leiden and Boston: Brill, 2014), see especially 198–9.

[124] "In einem komplizierten Prozeß von Adaptionen zwischen Rollenentwurf und individuellen Daten erfolgt die Ausbildung eines komplexen Autorschaftsbildes, das als stark stilisierte Selbstpräsentation des Verfassers angeboten wird," see Christel Meier, "Autorschaft im 12. Jahrhundert. Persönliche Identität und Rollenkonstrukt," in *Unverwechselbarkeit. Persönliche Identität und Identifikation in der vormodernen Gesellschaft*, ed. Peter von Moos, Norm und Struktur 23 (Cologne: Böhlau Verlag, 2004), 215.

[125] This is an important point throughout the book of van 't Spijker, *Fictions of the Inner Life. Religious Literature and Formation of the Self in Eleventh and Twelfth Centuries*.

constituted by the spectrum between fiction and forgery extend to letter collections,[126] such as Volmar's forged letter heading Hildegard of Bingen's *Epistolarium* (mentioned earlier),[127] Bernard of Clairvaux's meticulous revisions of the contents of his letters and his modelling on literary examples,[128] or, most notoriously, the letter exchange of Heloise and Abelard.[129] They may also be shown to apply for other documents such as charters,[130] the donation charter of Charlemagne for example, topic of discussion further down this thesis.[131]

The rise of fictionality has a bearing on collaborative authorship. It indicates that authorial role-playing can be deliberately modelled, and that one takes a risk in assuming that the distribution of roles as presupposed within the text reflects a historical reality. Collaborative authorship, as well, could be 'stylized,' or indeed, become a trope with a literary function.[132] Aside from having a factual core, Hildegard of Bingen and Elisabeth of Schönau's explicit references to secretaries participating in the composition of their texts need also be read as literary devices through which the visionaries sought to assert their authority and credibility. The question, consequently, is which of these functions —historical or fictional— needs to be taken most seriously. To what extent are Hildegard and Elisabeth exercising a widely recognized humility topos in asserting their deficiencies, and to what extent are figures such as Volmar or Ekbert effectively contributing behind the scene?[133] In Elisabeth's visions, the authorship of respectively masculine and feminine sections has been interpreted to give the visions their ordering principle, in other words, collaboration becomes a means of narrative structuring.[134] Finally, one may point out that the Heloise-Abelard collection's greatest narrative feat is that it *reads* like a dialogue between a man and a woman, meticulously building up toward ideals for spiritual life.

[126] Constable, "Forged Letters in the Middle Ages."

[127] Van Engen, "Letters and the Public Persona of Hildegard."

[128] Wim Verbaal, "Voicing your Voice: the Fiction of a Life. Early Twelfth-Century Letter Collections and the Case of Bernard of Clairvaux.," *Interfaces* 4 (2017): 103–24.

[129] Discussed in full from p. 235 onward. For the letters' fictional aspects, see Wim Verbaal, "Epistolary Voices and the Fiction of History," in *Medieval Letters – Between Fiction and Document*, ed. Christian Høgel and Elisabetta Bartoli, Utrecht Studies in Medieval Literacy 33 (Turnhout: Brepols, 2015), 1–23.

[130] Derek Pearsall, "Forging Truth in Medieval England," in *Cultures of Forgery: Making Nations, Making Selves*, ed. Judith Ryan and Alfred Thomas (New York and London: Routledge, 2003), 3–14.

[131] See chap. 7, pp. 203–234.

[132] Lynn Staley Johnson, "The Trope of the Scribe and the Question of Literary Authority in the Works of Julian of Norwich and Margery Kempe," *Speculum* 66, no. 4 (1991): 820–38.

[133] A matter more fully discussed in pp. 147–176 and pp. 147–176.

[134] Marie-Geneviève Grossel, "Voir des yeux du coeur, faire la vision parole: quelques remarques sur les visions d'Elisabeth de Schönau," in *Desir n'a repos. Hommage à Danielle Bohler*, ed. Florence Bouchet and Danièle James-Raoul, Eidôlon 115 (Bordeaux: Presses universitaires de Bordeaux, 2015), 367.

# 2

# Computational Stylistics

## 2.1 Definitions and Theory

Perhaps not to its advantage, computational stylistics as a method combines two rather contested and/or disfavoured ideas within cultural-literary studies. The predicate 'computational' implies a type of automated, positivistic measurability and objectivity (in a post-modern age of reading), and 'stylistics' implies a computational counterpart or adaptation of the practice of stylistics, which has been criticized due to its all too restrictive concentration on formal aspects of literature.[1] To make matters even more complicated, stylistics encapsulates another impossible concept to define: style. We will show here that these terms —computational, stylistics, style— might all be rather misleading when seeking to define or demarcate the field in which computational stylistics operates. In effect, not everyone will agree that computational stylistics' object of study corresponds to style at all.

### 2.1.1 Computational

#### *Etymologiae*

The Latin verb *computare* (of which our modern-day word 'computer' is derived),[2] is a compound lemma consisting of the prefix *com-* and the verb *putare* ('to reckon'). The prefix *com-* (*cum-*) denotes an intensivation of the verb that follows, in the sense

[1] Craig, when writing on this specific topic in Hugh Craig, "Authorial Attribution and Computational Stylistics: If You Can Tell Authors Apart, Have You Learned Anything About Them?," *Literary and Linguistic Computing* 14, no. 1 (1999): pointed out the essays by literary critic Stanley Fish, more specifically; Stanley E. Fish, "What Is Stylistics and Why Are They Saying Such Terrible Things About It?," in *Approaches to Poetics: Selected Papers from the English Institute*, ed. Seymour Chatman (New York: Columbia University Press, 1973), and; Stanley E. Fish, "What Is Stylistics and Why Are They Saying Such Terrible Things About It?—Part II," *boundary* 2, no. 1 (1979): 129–146.

[2] Our modern-day, unambiguous understanding of a computer being a machine that processes numerical data dates from around 1969. Before the invention of the machine, the word 'computer' existed to denote a 'person who calculated,' see Edita Karaliutė, "A Case of Bidirectional Metaphor: A Computer as a Human Being and the Reverse," *Jaunųjų Mokslininkų Darbai* 2, no. 23 (2009): 45.

that 'reckon' is iterated a number of times. Multiple interpretative tasks or 'reckonings' are lined up and ultimately brought together (*com-*). In the original sense, the Latin word therefore often meant little else than 'to count' or 'to calculate.' However, we will see that in discussing computational stylistics, the basal and context-devoid signification of "lining up and executing a variable number of *n* interpretative tasks" as it comes forward from the Latin verb is a much more transparent and helpful way to think about *computatio*. The true feat of the computer is its ability to perform *computatio* as in "intensive iteration of interpretation." Adhering to this etymological fundament of *computatio* makes one better understand why some computational stylisticians nowadays believe that their computational approach to texts comes with an unsurpassable advantage in opposition to traditional readings of texts and constitutes a true computational turn:[3] interpretation can be intensified through using the computer. Connections between texts that would take traditional close readers very long to discover —or that would not have been discovered at all— are uncovered through computational data analysis.

## Distant Reading and Macro-Analysis

This idea of enhanced analytical potential offered to the humanities scholar by means of computers has been the theoretical spearhead of literary computing. It was (and is) most famously exploited as such by the Stanford Literary Lab, notably by leading scholars such as Franco Moretti and Matthew L. Jockers. Indeed, Moretti's much-debated concept of distant reading[4] and Jockers's macroanalysis[5] both refer to a similar idea: quantitative approaches to literature have become not only feasible but necessary, as the literary scholar's data is growing expansive to the point of deluge and as the computational power to handle this excess of data is available for the first time in academic history. Moretti's and Jockers's idea is that computational methods are capable of detecting larger patterns within the fabric of literary texts which the ordinary human mind can impossibly process. The arrival of big data at the modern-day researcher's disposal has increased the need of a macro-scalar approach that reduces texts to their formal essence: "fewer elements, hence a sharper sense of their overall interconnection. Shapes, relations, structures. Forms. Models."[6] Moretti, who mainly enjoyed

[3] David M. Berry, "The Computational Turn: Thinking About the Digital Humanities," *Culture Machine* 12 (2011): 1–22.

[4] A collection of his most seminal papers, which arguably form the manifestos of distant reading, can now be found in Franco Moretti, *Distant Reading* (London: Verso, 2013); but Moretti introduced the approach as early as 2000, in Franco Moretti, "Conjectures on World Literature," *New Left Review* 1 (2000): 54–68.

[5] Matthew Lee Jockers, *Macroanalysis: Digital Methods and Literary History*, Topics in the Digital Humanities (Champaign, IL: University of Illinois Press, 2013).

[6] See the first page of his introductory chapter "Graphs, Maps, Trees," to a collection of three essays formerly published in *New Left Review* (*NLR*) in the course of 2003 and 2004 (*NLR* 24, 67–93; *NLR* 26, 79–103; *NLR* 28, 43–63), and presently collected in Franco Moretti, *Graphs, Maps, Trees. Abstract Models for Literary History* (New York: Verso, 2005).

a Marxist and formalist formation and exhibits a profound faith in the potential of the exact sciences, pleads for quantitative distance instead of qualitative intimacy with source material. Jänicke et al. have opposed close reading in its ability "to read the source text without dissolving its structure," broadly defined as "the thorough interpretation of a text passage by the determination of central themes and the analysis of their development." Distant reading does the opposite: "It aims to generate an abstract view by shifting from observing textual content to visualizing global features of a single or of multiple text(s)."[7]

Much (perhaps a little too much) can be interpreted from such 'big' terms as macroanalysis and distant reading, and Jockers and Moretti evidently had more than computational stylistics in mind when they entrusted their ideas to paper. Moretti said so himself: "The quantitative approach to literature can take several different forms —from computational stylistics to thematic databases, book history, and more."[8] The truth is that Moretti's and Jockers's works audaciously flirt with a dizzyingly diverse array of paradigms, disciplines and schools. Few analogies are shunned to bring the point home. Not only big data, but also big theories and their respective jargons uncomplicatedly converge, quite deliberately made to transcend all geographical and chronological boundaries (indeed: globalism and *longue durée*). Physics, mathematics, biology, geology, economics, historical studies, social sciences, linguistics: from Charles Darwin to Lucien Febvre to Alan Turing, ... Moretti's and Jockers's vision for literary studies affiliates very little with contemporary literary theory aside from the occasional whiff of Russian formalism (or "quantitative formalism," as they habitually term it).[9] Instead, they happily breathe in the fresh wind coming from the exact sciences as an alternative approach for future literary studies. In substitution of subjective and anecdotal approaches to literature, they propose rational, empirical and verifiable methods to validate the reading experience, and justify their methods by applying them to a spectacular range of different research topics: plot, tradition, gender, nationality, theme, influence, narratology, characterization, ... It is *all* there for the taking, and they take it all. Quite imaginably, the resulting books are not exactly modest, but they burst with ideas that have been and still are highly inspiring to the present-day field of DH. What typifies their theories (and the computational turn in general) is a sense of constructiveness, scientific optimism and firm faith in human and technological progress which literary theory abandoned decades ago. "A more rational literary history. That is the idea."[10]

The type of stylistics maintained in this book is indeed 'computational' (or, if you

[7] Stefan Jänicke et al., "On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges," in *Proceedings of the Eurographics Conference on Visualization (EuroVis)—STAR*s, ed. Rita Borgo, Fabio Ganovelli, and Ivan Viola (Cagliari: The Eurographics Association, May 2015), 84.

[8] Franco Moretti, "Graphs, Maps, Trees. Abstract Models for Literary History—1," *New Left Review* 24 (2003): 68.

[9] Sarah Allison et al., "Quantitative Formalism," in *Canon / Archive*, Studies in Quantitative Formalism from the Stanford Literary Lab (New York: n+1 Foundation, 2017), 1–31.

[10] Moretti, "Graphs, Maps, Trees. Abstract Models for Literary History—1," 68.

will, 'statistical'), and although one does not necessarily have to agree with all of Moretti's ideological viewpoints, computational stylistics can arguably be justified as a distant reading of style. Many of the principles which scholars such as Moretti and Jockers formulated as what they believe constitutes the impact of the computational in literary studies in fact applies. Computational stylistics indeed brings "computation into criticism,"[11] or belongs to what Ramsay came to call "algorithmic criticism."[12] I sum up a few reasons here to make my point:

- Computational stylistics implies an abstraction of the original text; or, it is modelled. The original text goes lost (i.e. its word order and symbolic appearance are abandoned) as it is transformed into a so-called 'bag of words'[13] and cast into a numerical shape or other representation of some kind.[14]
- Computational stylistics is not based on close reading, but departs from observations that are automatically extracted, i.e. data-driven.
- Computational stylistics can process a much larger amount of stylistic information in a shorter span of time than the human mind, and through its use of algorithms it can indeed track down patterns and connections which the human mind would not have been able to track down.[15]
- Computational stylistics in the state of the art still mostly relies on form, although experiments with grammatically, syntactically and semantically annotated text are increasingly gaining leeway.[16]
- Computational stylistics departs from a common ground that is both verifiable and replicable. The observation on record never changes, and may as such be termed empirically more objective.[17]

### 2.1.2 Stylistics

The second component of the methodology of this book suggests that computational stylistics is a branch of stylistics; a computational variant of the discipline that —

[11] I refer here to John F. Burrows, *Computation Into Criticism: A Study of Jane Austen's Novels and an Experiment in Method* (Oxford: Clarendon Press, 1987).

[12] Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism*, Topics in the Digital Humanities (Champaign, IL: University of Illinois Press, 2011).

[13] Yin Zhang, Rong Jin, and Zhi-Hua Zhou, "Understanding Bag-of-Words Model: A Statistical Framework," *International Journal of Machine Learning and Cybernetics* 1, no. 1 (2010): 43–52.

[14] We will return to this phenomenon, which falls under feature extraction, in 3.A.3.4 of this book.

[15] For instance, it is quite easy for the human mind to detect similarity or difference in *present* elements (i.e. one can easily compare two texts and spot which words those two texts have in common or not). The opposite, detecting similarity or difference in *absent* elements, is much harder and requires far more processing power (i.e. which words do these texts *not* contain or contain in a lesser degree so that it sets itself apart and compares as different against other styles at large)?

[16] For instance through more complex word representations such as word embeddings, see Tomas Mikolov et al., "Distributed Representations of Words and Phrases and their Compositionality," *Advances in Neural Information Processing Systems* 26 (2013): 3111–9.

[17] Of course, one can disagree on the transformation that preceded the initial observation, since the casting of a text into a numerical guise of choice always implies a bias from the analyst. Moreover, the leap from this initial observation to a meaningful interpretation is considerably less straightforward.

according to Leo Spitzer— "might bridge the gap between linguistics and literary history."[18] This is a complicated issue, as stylisticians have never univocally agreed on what exactly constitutes their practice, nor on what constitutes their object of study: style. It is apparent that the word is indeed used differently across various specialized fields, and brings along a variety of connotations.

## Style

One will frequently encounter the word 'style' used rather loosely without it being problematized, and not seldom the distinction between the term and other terms such as language, discourse, register or rhetorics remains unclear. The problems of defining style are legio. For one, style can be both personal as collective. Think for instance of specialist jargons, artistic movements, even Erich Auerbach's classical and Elohistic styles which he saw converging in the European Middle Ages.[19] In that same light one will observe that for some style is conceived as innate whereas for others it is nurtured, i.e. individually or collectively determined, or a combination of both. Some might believe the individual is always to be seen in cooperation and in imitation with a collective. But then again, an individual might also be surprisingly flexible in adapting his or her writing style. The surrealist Raymond Queneau's *Exercises de style*, a twentieth-century collection of the same story told over again ninety-nine times in a different style, springs to mind as just one obvious example.[20] One might encounter those who believe that style —as it did for Roman Jakobson (1896–1982) and other Russian formalists— is closely entwined with the poetic function of that text, i.e. its literary value. This raises the question if there is a distinction between practical and poetic language, and if a package leaflet has a style, for instance. More questions arise. Does proficiency in language correlate with stylistic flexibility? Is style a regional or language-bound phenomenon, or is there something as cross-linguistic and translatorial styles?[21] Is style an evolutionary phenomenon, where certain styles dominate and extinguish others?[22] How does style relate with nationality, gender or ethnicity? Should we delimit style to language alone, or does it also pertain to *mise en page* (not only layout but also other elements of page decoration), to typography and to its material carrier?

[18] "Stylistics, I thought, might bridge the gap between linguistics and literary history," see Leo Spitzer, *Linguistics and Literary History: Essays in Stylistics*, 2nd ed. (New Jersey: Princeton University Press, 1967 (1948)), 11.

[19] Erich Auerbach, *Mimesis: Dargestellte Wirklichkeit in der abendländischen Literatur*, 1st ed. (Bern: A. Francke A. G., 1946).

[20] Raymond Queneau, *Exercices de style*, Gallimard (Paris, 1997 (1947)).

[21] See for instance John Burrows's work on translations of Juvenal in John F. Burrows, "The Englishing of Juvenal: Computational Stylistics and Translated Texts," *Style* 36, no. 4 (2002): 677–750.

[22] Franco Moretti, "The Slaughterhouse of Literature," *Modern Language Quarterly* 61, no. 1 (2000): 207–27.

## Stylistics

When it comes to the similarities and differences between computational stylistics and stylistics as such, a few points should be emphasized. One can say that already as early as the structuralist movement with scholars such as Jakobson, stylistics had, due to its adoption of the linguist's drive to observe grand movements in language and explain them, left the door open for systematic, logical and mathematical principles underlying language.[23] In a sense, stylistics always has had computational sympathies, and has always flirted with the possibilities provided by the exact sciences. Both fields —stylistics and computational stylistics— consequently share an aspiration for the scientificity of literary studies (I leave open whether or not such an aspiration is realistic).[24] They also care more for the form of language and how that form relates to meaning, rather than for the post-structuralist problematization between that form and meaning that arose in the late 1960s.

Nevertheless, despite the kinship suggested by the field's names, stylistics and computational stylistics have grown out to be two quite distinctive areas of research. Some might find this problematic.[25] The first is, for instance, far more committed to linguistics. Stylistics still relies on a close, careful reading, and arms itself against subjectivity by founding their literary readings upon linguistic ground rules. In the field of computational stylistics, such a close engagement with the source text is either absent or else builds upon a former quantitative reading. Its actual association with linguistics is somewhat decreased, in the sense that its engagement with subdisciplines such as phonetics, grammar, morphology, syntax, etc. has been traded in for an arsenal of statistical techniques that provide their scientific tools. Moreover, instead of fostering an interest for the aesthetic effect of the text, computational stylistics' subtasks are often more positivistic and triggered by historical and forensic-attributive questions of authorship and text provenance.[26] Finally, one could also argue that stylistics, in attempting to discover the function of a text, places more emphasis on finding linguistic rule violation[27] or departure from the norm causing defamiliarization, which stands

[23] Roman Jakobson, ed., *Structure of Language and Its Mathematical Aspects*, vol. 12, Proceedings of Symposia in Applied Mathematics (New York: American Mathematical Society, 1961).

[24] The skeptics might want to read Stanley Fish's critiques on stylistics. See n. 1 of this chapter. Computational stylistics has also had its share of critique, see Joseph Rudman, "The State of Authorship Attribution Studies: Some Problems and Solutions," *Computers and the Humanities* 31 (1998): 351–65; and Joseph Rudman, "The State of Non-Traditional Authorship Attribution Studies—2012: Some Problems and Solutions," *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 259–74.

[25] In arguing what the ideal computational stylistician should bring to the study, Joseph Rudman demanded that "A strong grasp of [stylistics] is necessary," see Joseph Rudman, "Non-Traditional Authorship Attribution Studies in the *Historia Augusta*," *Literary and Linguistic Computing* 13, no. 3 (1998): 151.

[26] Although one is of course expected to make the distinction between stylistics and authorship studies, see Hugh Craig, "Stylistic Analysis and Authorship Studies," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth, Blackwell Companions to Literature and Culture (Oxford, UK: Blackwell, 2004), 273–88.

[27] Henry G. Widdowson, *Stylistics and the Teaching of Literature*, Applied Linguistics and Language Study (New York: Routledge, 2013 (1975)), 17.

in contrast to computational stylistics' emphasis on the opposite, i.e. discovering the rules to which the text complies (after which, of course, the outliers are analyzed). The different instruments by which each of these stylistics approach style again triggers the question of what exactly each of them is measuring, and if we can be certain that this object of interest —if it is style, indeed— coincides at all.

In effect, defining style is a very complex and thorny pursuit, and the question whether or not computational stylistics indeed automatically culls elements from a text that can rightly be called stylistic as such may be a false question. Its answer must automatically rely on convention and personal conviction. Although I find it valuable to keep track of the polysemy of the term 'style,' others have done this better and more fully than I can here.[28] Perhaps a more useful question is how one expects computational stylistics —a distant reading of style— to contribute to the understanding of this mystery that is style and language? What can a computational approach teach us about style that we did not already know? And, more centrally to this thesis's main focus: how does computational stylistics' definition of style relate to matters of authorship and authorship attribution?

### 2.1.3 Style and Attribution

When it comes to authorial style, computational stylistics has at least one revolutionary find to offer, and that is the observation that it is possible to statistically itemize —or formally trace within a text— features that are unique or particular to an individual's writing habits. There is a unique pattern to how a human writes, a kind of stylistic DNA, if you will, an authorial fingerprint which one may also find referred to as a 'stylome.'[29] Whether or not a stylome exists has been tested and proven in numerous experiments, to the extent that it has become possible to assert that there are aspects of personal style susceptible to empirical measurement.[30] In the definition of computational stylistics, personal style is encrypted —or imprinted— in a text through complex, multivariate patterns, and the computational age disposes analysts of the analytical power to decrypt it in full and expose it to the human eye. Correspondingly, this presents computational stylisticians with the ability to segregate documents of varying authorship, and, if possible, attribute the measured pattern to an identifiable author. An important returning argument is that style, as it is detected by computational methods, is largely a subconsciously produced given. There are aspects to writing style which authors do not have full control over, because they are invisibly lost in larger, abstract trends which even the best imitator will not be able to

[28] See for instance the article of J. Berenike Herrmann, Karina van Dalen-Oskam, and Christof Schöch, "Revisiting Style: a Key Concept in Literary Studies," *Journal of Literary Theory* 9, no. 1 (2015): which explicitly seeks out the confronting definitions of style in both traditional as digital, non-traditional humanities fields.

[29] Don Foster, *Author Unknown. On the Trail of Anonymous* (New York: Henry Holt / Company, 2000), 12.

[30] van Halteren et al., "New Machine Learning Methods Demonstrate the Existence of a Human Stylome."

access.

Taking into account the preceding observations, we have arrived at the following working definition: computational stylistics is the discipline of detecting statistically traceable and verifiable aspects of a document which an intimate appreciation of a document cannot provide, with the purpose of arriving at the identification of the document and/or an objectively founded assessment of that document's formal properties.

## 2.2 The History of Computational Stylistics

### 2.2.1 Forefathers: The Nineteenth Century

#### Mendenhall and 'Curves of Composition'

In spite of what the often encountered discourse surrounding computational stylistics suggests with such adjectives as 'revolutionary' and 'innovative,' many aspects of the research field are arguably "as old as the hills."[31] Well before the age of computing, in pursuing a casual suggestion by mathematician Augustus De Morgan (1806–1871),[32] Thomas C. Mendenhall (1841–1924) in the late nineteenth century published an article in *Science* on what he called "characteristic curves of composition."[33] Originally being a physicist and meteorologist, Mendenhall was fascinated and inspired by electromagnetic radiation and light wave-length, and took up Charles Dickens's novels to confirm if such natural laws apply elsewhere as well. He manually started counting words and their lengths for very small text samples in such novels as *Oliver Twist* and *Great Expectations* (1,000 words).[34] Consequently, he plotted their word-length on the x-axis and relative frequency on the y-axis. The visual evidence suggested that authors also seemed to exhibit the wavelengths Mendenhall was fascinated by:

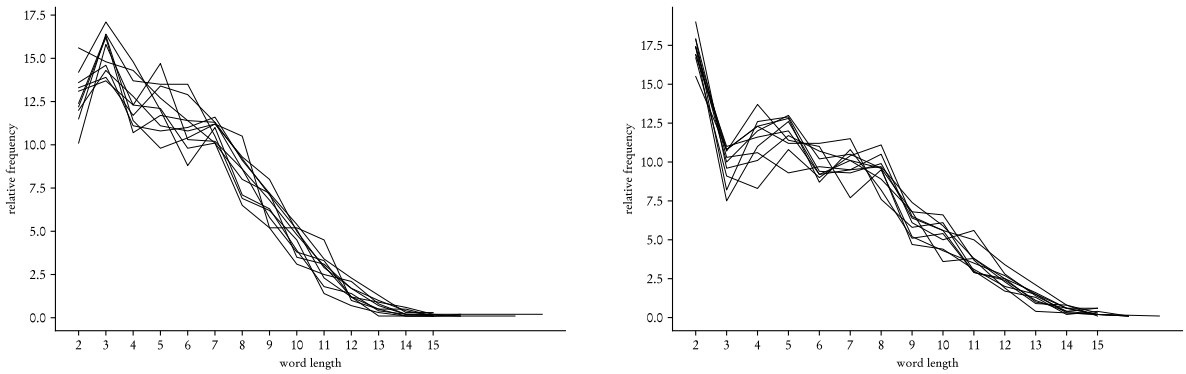
> The validity of the method as a test of authorship, then, implies the following assumptions: that every writer makes use of a vocabulary which is peculiar to himself, and the character of which does not materially change from year to year during his productive period; that, in the use of that vocabulary in composition, personal peculiarities

[31] Javier Calle-Martín and Antonio Miranda-García, "Stylometry and Authorship Attribution: Introduction to the Special Issue," *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 251.

[32] De Morgan had made the suggestion in a letter to a friend, that the average length of words in the Pauline epistles (a classic example of authorship dispute), might well present leads to uncovering knowledge regarding their authorship, or even settle the matter.

[33] Thomas Corwin Mendenhall, "The Characteristic Curves of Composition," *Science* 9, no. 214 (1887): 237–49. Around the same time, William Benjamin Smith (1850–1934) under the penname of Conrad Mascol, published two articles on the authorship of the Greek Pauline epistles using similar methods, see Conrad Mascol, "Curves of Pauline and Pseudo-Pauline Style I," *Unitarian Review* 30 (1888): 452–60; and "Curves of Pauline and Pseudo-Pauline Style II," *Unitarian Review* 30 (1888): 539–46.

[34] We will further elaborate on text sampling and segmentation in the next chapter from p. 91 onward. For now it suffices to say that a sample is a segment of text (measured by a number of words), and the general presupposition is that the longer this sample, the more certainty can be gathered on the sample's stylistic content and authorship.




(a) Bernard of Clairvaux (b) Hildegard of Bingen

Figure 2.1: Mendenhall curves for the first 10,000 words of Bernard of Clairvaux's *Sermones super Cantica Canticorum* and Hildegard of Bingen's *Liber diuinorum operum* (text samples are 1,000 words). The x-axis denotes word length, the y-axis denotes the frequency by which words with that length occur. Punctuation was deleted and all characters were tokenized in lowercase.

> in the construction of the sentences will, *in the long-run*, recur with such regularity that short words, long words, and words of medium length, will occur with definite relative frequencies.[35]

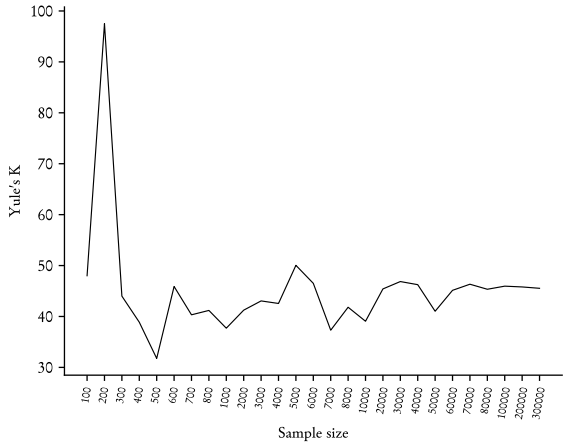
Today we know that the Mendenhall method is not reliable for accurately distinguishing between works of different authorship. The method has become computationally easy, and requires but a few minutes of coding as opposed to Mendenhall's intensive manual counting. By technological innovations and the arrival of multivariate analysis and machine learning (more elaborately explained on p. 322) we can now assess with a higher degree of reliability the results originally yielded by Mendenhall's methods, and they certainly are flawed if one's objective is to tell apart documents written by different authors. Specifically in Mendenhall's later work on Shakespeare plays,[36] it became all too obvious that he payed too slight attention to what other factors might have contributed to the differences he measured, such as genre.[37] It is also clear, from a contemporary perspective, that Mendenhall's technique tapped into word frequency rather than word length by using this method. Shorter words automatically occur more frequently, because they often belong to the class of function words, e.g. conjunctions, pronouns, prepositions, adverbs, particles, etc. (see below, p. 67).

Nevertheless, irrespective of the method's outdatedness, one is able to see that Mendenhall had bumped into an intriguing matter all the same: a single numeric function worked particularly well to distinguish between works by different authors. For

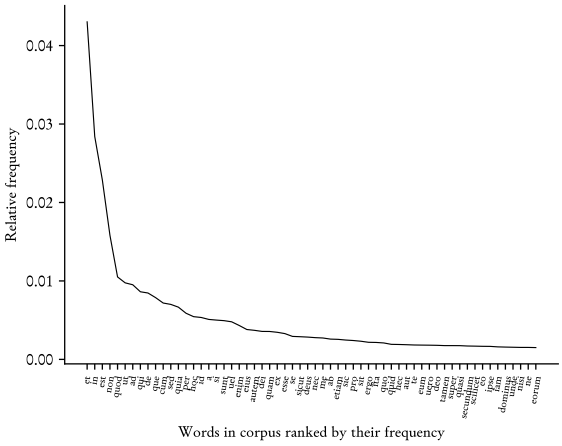
[35] Thomas Corwin Mendenhall, "The Characteristic Curves of Composition," *Science* 9, no. 214 (1887): 238–9.

[36] Thomas Corwin Mendenhall, "A Mechanical Solution of a Literary Problem," *Popular Science Monthly* 60 (1901): 97–105.

[37] This criticism was first voiced in Carrington B. Williams, "Mendenhall's Studies of Word-Length Distribution in the Works of Shakespeare and Bacon," *Biometrika* 62, no. 1 (1975): 207–212.




(a) Yule's *K*. The higher *K*, the less rich the vocabulary. Note the striking constancy of *K* despite the differing sample length.




(b) Zipf's Law. The 60 most common words' relative frequencies are plotted on the y-axis.

Figure 2.2: Intuition of G. Udny Yule's constancy measure *K* and George Kingsley Zipf's law of relative frequency. Yule's lexical richness was executed on a 1.3 million words corpus of all the texts of Bernard of Clairvaux. Zipf's law was executed on the main and benchmark corpus added together, resulting in some 9.7 million words. For inspection of the underlying data, see p. 308.

instance, when observing the general trend of the Bernard of Clairvaux's Mendenhall curve as opposed to that of Hildegard of Bingen in fig. 2.1 (different curves of different text samples are plotted on top of one other), one notices that each of their curves have similar author-specific trends, which sets each of their respective curves apart from one another. One can spot, for instance, that Hildegard uses many more small words consisting of two characters (the function word *in* plays a large role in this),[38] which causes a somewhat more dramatic drop in frequency of longer words (three to four characters). Bernard of Clairvaux tends to use longer words more often, and even starts with two-character-words lagging behind on the longer words.

## Yule and 'Lexical Richness'

Other notable scholars in the first half of the twentieth century followed Mendenhall's tracks in their own quest of finding statistical laws to which natural language obeys. One of them was the British statistician G. Udny Yule (1871–1951), who first proposed mean sentence length instead of word length as a feature in which statistical undercurrents were at work. Armed with empirical deductions he verified suspicions that the *De imitatione Christi*, the influential and intensively translated medieval handbook of the Modern Devotion movement, was written by the fifteenth-century

[38] "Given the visionary discourse developed in much of her writings —even in her letters— it is not surprising to come across an intensive use of the preposition *in* in Hildegard's letters. She repeatedly sees things *in* divine visions; she continuously searches the allegorical meanings buried *in* the multitude of details that she discovers in her visions," see Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 209.

Augustinian cleric Thomas à Kempis. The work's authorship had for long been contested and instead attributed to Jean (Charlier) de Gerson[39] by such scholars as the Sicilian Benedictine Constantino Cajetan (1560–1650).[40] Yule further elaborated his findings in a seminal study in 1944, in which he introduced an additional measure, the *K* constancy measure (fig. 2.2a).[41]

Yule's *K*, as it came to be called, is a statistical vocabulary richness measure that is —theoretically— invariant with text length.[42] Especially the latter condition, that vocabulary richness or lexical diversity can be measured irrespective of a text's length, is not self-evident. An intuitive reflex namely would be that —if we want to measure the lexical diversity of such a sample— it will be hard to compare the vocabulary richness of shorter samples with longer samples, since short text automatically has less space and vocabulary to treat a large range of topics than long text. One can observe in fig. 2.2a, however, that from a sample length of about 500 words onwards already, Yule's *K* succeeds at finding a constant measure floating around 45.13 that describes the lexical diversity of Bernard of Clairvaux. As the sample sizes increase on the y-axis, the lexical diversity of Bernard of Clairvaux can possibly be described by a singular, constant ratio. A lifeline with a slight pulse. Yule's goal was to define what vocabulary (lexical preference) constituted *as a whole*, which he found superior to subjective, anecdotic techniques of determining authorship. Or, to say it in Yule's own charming words: "To tell me that there is a small mole on Miranda's cheek may help me to identify the lady, and may in conceivable circumstances be quite useful information to the police, but it hardly amounts to a description of her alluring features."[43]

## Zipf and the 'Principle of Least Effort'

Another scholar worth mentioning here —who was active around the same period as Yule— is the American linguist and philologist George Kingsley Zipf (1902–50). Zipf was driven by similar empirical observations of regularity within human language, specifically to the works of James Joyce and to Chinese phonology.[44] The latter two subjects might appear somewhat remote from each other, but as it turns out Zipf's goal was directed toward explaining rules of human behaviour and natural phe-

[39] Yule, "On Sentence-Length," 372ff.

[40] Not coincidentally in the sixteenth century, when the book reached the peak of its popularity, and was intensively translated in the vernacular. For Cajetan's attribution, see Thomas Joseph Shahan, "Constantino Cajetan," in *The Catholic Encyclopedia*, ed. Charles G. Herbermann, vol. 3 (New York: Robert Appleton, 1913), 360–1.

[41] George Udny Yule, *The Statistical Study of Literary Vocabulary* (New York: Archon Books, 1968 (1944)).

[42] Although Yule proposed the measure as one by which to distinguish documents of different authorship, this is in practice never guaranteed, see Fiona J. Tweedie and R. Harald Baayen, "How Variable May a Constant Be? Measures of Lexical Richness in Perspective," *Computers and the Humanities* 32 (1998): 323–52.

[43] Yule, *The Statistical Study of Literary Vocabulary*, 2.

[44] George Kingsley Zipf, *Selected Studies of the Principle of Relative Frequency in Language* (Cambridge, MA: Harvard University Press, 1932).

nomena in general. Natural language, so Zipf contended, was but a glimpse of a larger, universal trend which —two years before his death— he theoretically developed into "the principle of least effort," which obeys to only this rule of thumb: "the entire behavior of an individual is at all times motivated by the urge to minimize effort."[45]

> The study of words offers a key to an understanding of the entire speech process, while the study of the entire speech process offers a key to an understanding of the personality and of the entire field of biosocial dynamics.[46]

The principle of least effort was Zipf's attempt to explain the equation known today as Zipf's law: $f(r) \propto \frac{1}{r^a}$. The law stated that the frequency ($f$) of a word has a frequency inversely proportional to its frequency rank ($r$),[47] where the most frequent word will occur approximately twice as many times as the second most frequent word, three times as the third most frequent word, and so on. That Zipf's law also applies to a medieval Latin corpus can be observed in fig. 2.2b, where the frequencies of the 60 most common words of the main and benchmark corpus added together approximately follow a Zipfian power law curve.[48]

For Zipf, the principle of least effort explains this phenomenon: a writer chooses the path of least resistance and highest efficiency, and draws from a fixed lexical resource in order to obtain an objective with the least amount of effort. Especially short words with few characters demand little storage. Reading Zipf is a dizzyingly estranging experience, especially when after some chapters even Freud's dream worlds are demonstrated to obey to the Zipfian logic.[49] The true underlying cause of the empirical phenomenon that has been termed Zipf's law is still topic of debate.[50]

### 2.2.2 The *Federalist Papers* and Function Words

That computational stylistics has its far roots in the nineteenth century and the early twentieth century with scholars such as Mendenhall, Yule and Zipf should not be surprising. That its breakthrough came several years before Roland Barthes proclaimed the author dead, on the other hand, is a fine example of historical irony.[51] It was not until the 1960s that computational analysis of style earned serious consideration by

[45] George Kingsley Zipf, *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology* (Cambridge, MA: Addison-Wesley Press, Inc., 1949), 3.

[46] Ibid., 19.

[47] Steven T. Piantadosi, "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions," *Psychon Bull Rev* 21 (2014): 1112–30; Zipf, *Selected Studies*.

[48] In fact, Zipf's law applied to many other populations, Eric S. Wheeler, "Zipf's Law and Why It Works Everywhere," *Glottometrics* 4 (2002): 45–8.

[49] There is something psychotic about Zipf's work. See the eight chapter entitled "The Language of Dreams and of Art," in Zipf, *Human Behavior and the Principle of Least Effort*, 313ff.

[50] Wheeler tested Zipf's law for random numbers and concluded that "if such a pattern happens when the subject matter is random and content-free, then the pattern is the result of the method of analyzing and presenting the data, and not a property of the subject matter at all," see Wheeler, "Zipf's Law and Why It Works Everywhere," 45.

[51] Barthes, "The Death of the Author."

the academic world. In 1964, Frederick Mosteller and David L. Wallace, two American statisticians, "blunder[ed] into an historical and literary controversy,"[52] a blunder which along the way so happened to become the founding work for non-traditional authorship attribution.

In their work, which was also the start of a multivariate analysis of style, Mosteller and Wallace were able to demonstrate through Bayesian methods[53] not only that it is possible to statistically determine formal features in our language which can be aggregated and statistically analyzed, but that especially the frequencies of a set of small, insignificant yet common markers in English such as 'enough,' 'whilst,' 'upon' and 'the' provided valuable lexical information for a statistician to make highly accurate predictions concerning a text's authorship. These marker words, which often belong to the top most frequent words stratum of a text, enabled them to arrive at conclusions concerning the attribution of some of the pseudonymous late eighteenth-century *Federalist papers*, 146 relatively short pseudonymous articles to ratify the constitution.[54]

The marker words which Mosteller and Wallace spoke of were words predominantly known in the literature as function words. These are adverbs, conjunctions, prepositions, pronouns, articles, auxiliary verbs, ... they represent the syntactic and grammatical stuffing of a text which one passes over because they have little to no semantic content, and are therefore considered topic-independent. Authors, moreover, have less control over the rhythm and frequency by which they apply them. They often go unnoticed for authors and readers, which implies that imitators have decreased chances of concealing their tracks when attempting to assimilate with a target text or author. Today, still, despite the numerous experiments with different features and techniques, many practitioners of computational stylistics believe that they are very robust predictors when it comes to segregating writing styles.[55]

Without going into the technical details yet of how such plots are devised,[56] and shown here purely from an illustrative point of view, fig. 2.3 gives an idea of the power of function words. The idea is that both Bernard's *Sermones super Cantica Canticorum* and Hildegard's *Liber divinorum operum* were segmented into what is known in the literature as text samples.[57] Each text sample is a data point in the figure. As one can see, text samples of same authorship (here coloured for the sake of convenience in yellow —Bernard— and green —Hildegard) cluster together. The algorithm was, obviously, uninformed in advance which text samples belonged to

[52] Frederick Mosteller and David L. Wallace, *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*, 2nd ed., Springer Series in Statistics (New York: Springer Verlag, 1984 (1964)), 1.

[53] We return to Naive Bayes in A.3.5 on p. 324.

[54] The political articles were written in the wake of the American revolution (1770s and 1780s). Seventy-seven were published in newspapers and twelve were either in dispute between Alexander Hamilton and James Madison, or were regarded as co-authored.

[55] Kestemont, "Function Words in Authorship Attribution."

[56] One can jump ahead to A.3.6 on p. 348 to learn more about principal components analysis (PCA).

[57] Sampling is treated in the next chapter on p. 91.

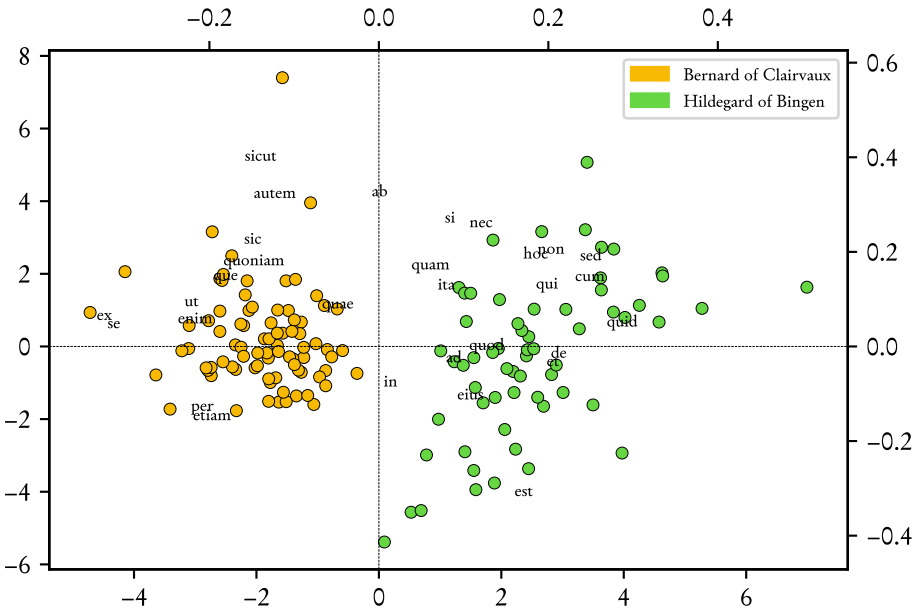



Figure 2.3: PCA cluster plot for Bernard of Clairvaux's *Sermones super Cantica Canticorum* and Hildegard of Bingen's *Liber divinorum operum*. The plot visualizes how easily one can separate authorial styles on the basis of 30 insignificant, small function words, which are plotted over the data points. Explained variance: 29.72%. Number of words: 30. Sample size: 2,000 *w*.

which author.

Although Mosteller and Wallace popularized the application of function words as predictors for authorship, there were others who drew attention to them before their time. For medieval Latin specifically, the scholar Lorenzo Minio-Paluello deserves mentioning. Minio-Paluello explored the usefulness of particles in segregating different translators of Aristotle's texts in the twelfth century.[58]

The application of function words was picked up and further refined by notable scholars such as John F. Burrows. Burrows applied his well-known Delta (Δ) method to the most frequent words of English texts from the Restoration period, and achieved very accurate results in distinguishing works of varying authorship.[59] Ever since Burrows, it has mainly been the engine of stylometric experiments that has received its fair share of renovation and improvement, whereas the fuel on which the engine runs has essentially remained unaltered: computational methods become increasingly accurate due to technological development, but very often this changes very little on a qualitative level. If one were to set out to detect meaningful differences between

[58] Lorenzo Minio-Paluello, "Iacobus Veneticus Grecus: Canonist and Translator of Aristotle," *Traditio* 8 (1952): see especially 280: "James shows some preference in the choice between words frequently used and of similar meanings : '*sicut*' is used five times in the course of the 'advice' but there is no '*quemadmodum* ,' '*tamquam*,' '*ut*,' '*velut*'; there are four '*sic*,' not one '*ita*'; four '*autem*' and two '*vero*'; four 'enim,' one '*namque*,' no '*nam*,' no '*etenim*'; three '*igitur*,' no '*ergo*,' no '*itaque*'; one '*huiusmodi*,' no '*talis*.'"

[59] Burrows and his Delta method (Δ) have a proper subsection in the appendix, see A.3.5 on p. 330. For Delta, see John F. Burrows, "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship," *Literary and Linguistic Computing* 17, no. 3 (2002): 267–87.

| 1-5 | 6-10 | 11-15 | 16-20 | 21–25 | 26–30 |
|---|---|---|---|---|---|
| et | que | sed | hoc | enim | sic |
| in | quod | per | nec | quae | sicut |
| est | qui | de | etiam | autem | ab |
| non | cum | se | quam | quoniam | eius |
| ad | ut | si | ita | ex | quid |

Table 2.1: Top 30 most frequent function words applied to distinguish between Hildegard of Bingen's *Liber Divinorum Operum* and Bernard of Clairvaux's *Super Cantica Canticorum*, as visualized in fig. 2.3. The top stratum of a text largely corresponds to a collection of insignificant, small function words, which are seldom conjugated or inflected.

authors' styles, most frequent words (MFW), and function words belonging to this group are arguably still the best state-of-the-art features from a computational perspective. Although attention has gone out to additional strata of lexical items (again by Burrows, and considerably improved by Hugh Craig),[60] and although (automatically) parsed syntactical metadata or grammatically-morphologically tagged (or annotated) categories have been proposed as additional, more sophisticated or deep features, they in theory *can*, yet seldom as a rule return improved accuracy.[61] The technical aspects of feature extraction and feature selection are further discussed in A.3.4 (p. 320).

It has been hypothesized that for analytic languages such as English and German, the function word theory works somewhat better than for synthetic (or agglutinative) languages with a high degree of inflection, such as Polish or —more central to our current interest— medieval Latin. The nominal cases can indeed to a large extent replace the role of function words, such as prepositions, which could raise skepticism as to the validity of the method. Rybicki and Eder have moreover argued by experimental evidence that this concern is not unjustified: different languages indeed yield different most-frequent-words strata and correspondingly yield slightly different results.[62] The granularity for lexical items is increasingly compromised for the most common words in agglutinative languages: much of the neighbouring textual content has become en-

[60] Referred to as Burrows' 'Zeta' —the middle of the word frequency spectrum— and 'Iota' —the lower strata of the word frequency spectrum—; see John F. Burrows, "All the Way Through: Testing for Authorship in Different Frequency Strata," *Literary and Linguistic Computing* 22, no. 1 (2007): 30; and Hugh Craig's variant that goes under the name of "Craig's Zeta," which implies making "two axes of authorial differentiation," with "words that are unusually common in the author of interest, and of hundreds that are not," see Hugh Craig and Arthur F. Kinney, *Shakespeare, Computers, and the Mystery of Authorship* (Cambridge, UK: Cambridge University Press, 2009), 18.

[61] In this light, Juola has warned for the "downside of such processing, especially for POS taggers,"since it might bring along "the introduction of errors in the processing itself," in their own turn "muddying the inferential waters." Patrick Juola, "Authorship Attribution," *Foundations and Trends in Information Retrieval* 1, no. 3 (2006): 265; Tempestt et al. argue that "though language-dependent, syntactic information is highly reliable assuming the availability of accurate and robust tokenizers, parsers, and part-of-speech taggers. Noise is introduced, however, when these tools are outdated," see Neal Tempestt et al., "Surveying Stylometry Techniques and Applications," *ACM Computing Surveys* 50, no. 6 (2017): 86:9.

[62] Jan Rybicki and Maciej Eder, "Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words?," *Literary and Linguistic Computing* 26, no. 3 (2011): 315–21.

capsulated in the items' morphemes (especially in suffixes, prefixes and case endings). We will indicate, however, through a state of the art of succesful applications in subsection 2.3, and through our own experiments in the next chapter, that even though the evaluative differences between languages stands as undeniable, function word frequencies —when treated with care— are still very reliable markers in Latin literature for distinguishing texts by different authors.

**Personality and Style**

Due to their potential to signal latent preferences of style and beyond, function word distributions have caught the eye of psycholinguists and social psychologists, who have studied the words' psychological dimension in relation with personal trauma and memory. Along an analogous trail as scholars such as George K. Zipf, who had played with Freudian ideas and a potential relationship between statistical regularities and the subconscious,[63] scholars such as James W. Pennebaker in *The Secret Life of Pronouns* have emphasized that these common yet inconspicuous markers often tell a lot more about the profile and mental state of their writers than taken at first glance.[64] Pennebaker's intention was to make the leap from the empirical observation of function word frequency distributions to "psychological profiling."[65] So Pennebaker demonstrates through experiments with function words on contemporary letters, poems, books, blogs, Tweets, conversations, and other texts that, for instance, the language of suicidal poets would contain more references to the first person singular ('I,' 'me,' 'myself'),[66] or that women are more prone to use personal pronouns and 'certainty words.'[67] Whether or not such psychological explanations can be transferred to text types in which the heart is not carried on the sleeve —such as medieval texts— is only the question.[68]

## 2.3 Computational Stylistics for Latin

'Latin' is a curiously artificial criterium for grouping a number of case studies, as sometimes the only binding aspect is linguistic (and even then one might be wary). In reality, Latinity comprises an extensive range of texts written over a long time span (predominantly from antiquity to the early 1800s), across modern borders in

[63] In the sixth chapter of his *Human Behavior*, Zipf attempted to find an objective definition of personality by taking language as a personality's underlying structure, see Zipf, *Human Behavior and the Principle of Least Effort*, 210ff.

[64] Pennebaker, *The Secret Life of Pronouns*.

[65] For a very early attempt, see Kim Luyckx and Walter Daelemans, "Using Syntactic Features to Predict Author Personality from Text," in *Digital Humanities 2008: Conference Abstracts* (Oulu, Finland: English Philology, June 2008), 146–9; and further reading in John Noecker, Michael Ryan, and Patrick Juola, "Psychological Profiling Through Textual Analysis," *Literary and Linguistic Computing* 28, no. 3 (2013): 382–7.

[66] Pennebaker, *The Secret Life of Pronouns*, 229.

[67] Ibid., 99.

[68] As beautifully problematized by Leclercq in Leclercq, "Modern Psychology."

divergent historical circumstances. Stylometric work has been done for Julius Caesar (100–44 BC),[69] Pliny the Younger (61–113 AD)[70] and Apuleius (c. 124–c. 170 AD),[71] as well as for Dante Alighieri's (1265–1321) *Epistolae*,[72] or John Milton's (1608–74) *De doctrina Christiana*.[73] At the same time, Latin's fossilized nature renders it surprisingly stable and robust in form, which has allowed to make meaningful analyses of Latin style for a much larger period than is feasible for modern natural languages:[74] one can marvel at the fact that it is reasonable to compare a Latin text by Alcuin (735–804) to a Latin text by Francesco Petrarch (1304–1374),[75] whereas comparing *Beowulf* to Geoffrey Chaucer's (1343–1400) *Canterbury Tales* —although both sets consist of two texts approximately at an equal chronological distance from one another— would come with almost unsurpassable problems of lexical and orthographical incompatibility.

### 2.3.1 The *Historia Augusta*

An important case study for the development of stylometric research in Latin, the "*Federalist Papers* for Latin stylometry," if you will, has been the attribution of the thirty biographies of second–third-century Roman emperors collected under the title of the *Historia Augusta* (hereafter *HA*). The texts themselves suggest to have been written by six distinct authors, but in 1889 Hermann Dessau tried to convince the scholarly world of his thesis that the *HA* is in fact written by no more than one single author at the end of the 4th century.[76] The series of articles following up a century after Dessau's thesis, in which statistics were applied to prove or disprove

[69] Mike Kestemont et al., "Authenticating the Writings of Julius Caesar," *Expert Systems with Applications* 63 (2016): 86–96; and some time later also: Olivia R. Zhang, Trevor Cohen, and Scott McGill, "Did Gaius Julius Caesar Write *De Bello Hispaniensi*? A Computational Study of Latin Classics," *Human IT* 14, no. 1 (2018): 28–58.

[70] Enrico Tuccinardi, "An Application of a Profile-Based Method for Authorship Verification: Investigating the Authenticity of Pliny the Younger's Letter to Trajan Concerning the Christians," *Digital Scholarship in the Humanities* 32, no. 2 (2017): 435–47.

[71] See Justin A. Stover and Mike Kestemont, "Reassessing the Apuleian Corpus: A Computational Approach to Authenticity," *The Classical Quarterly* 66, no. 2 (2016): 645–72; and Justin A. Stover et al., "Computational Authorship Verification Method Attributes a New Work to a Major 2nd Century African Author," *Journal of the Association for Information Science and Technology* 67, no. 1 (2016): 239–42.

[72] Downey et al. argued that Dante's Letter 13, or "Epistola a Can Grande,"is too untypical of Dante's style to be his, see Sarah Downey et al., "Lexomic Analysis of Medieval Latin Texts," *The Journal of Medieval Latin* 24 (2014): 225–275.

[73] Fiona J. Tweedie, David I. Holmes, and Thomas N. Corns, "The Provenance of *De Doctrina Christiana*, attributed to John Milton: A Statistical Investigation," *Literary and Linguistic Computing* 13, no. 2 (1998): 77–87.

[74] Francesco Stella, "Generic Constants and Chronological Variations in Statistical Linguistics on Latin Epistolography," in *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, ed. Tara Andrews and Caroline Macé, Lectio 1 (Turnhout: Brepols, 2014), 159–79; Maciej Eder, "Taking Stylometry to the Limits: Benchmark Study on 5,281 Texts from *Patrologia Latina*," in *Digital Humanities 2015: Conference Abstracts* (Sydney, July 2015), 1919–1924; Maciej Eder, "A Bird's Eye View of Early Modern Latin: Distant Reading, Network Analysis and Style Variation'," in *Early Modern Studies after the Digital Turn*, ed. Laura Estill, Diane K. Jakacki, and Michael Ullyot, vol. 502. Medieval and Renaissance Texts and Studies, New Technologies in Medieval and Renaissance Studies 6 (Toronto: Iter Press, 2016).

[75] Stella, "Generic Constants and Chronological Variations."

[76] Hermann Dessau, "Über Zeit und Persönlichkeit der *Scriptores Historiae Augustae*," *Hermes* 24, no. 3 (1889): 337–92.

him, are notoriously well-known for Latinists and computational stylisticians, not in the least because some of them present a slight embarrassment to the field. The very first in line to confirm Dessau's assumption of single authorship was Ian Marriott,[77] but unfortunately the latter relied on the dubitable feature of sentence length to bring his point home (the number of words per strong stop). What followed was a series of articles that not necessarily contradicted Marriott's belief in the *HA*'s homogeneity, but instead expressed harsh criticism towards Marriott's reliance on sentence length to make this assertion: "[...] Any statistical study carried out on an ancient text that is based upon the sentence runs the risk of examining, not the style of the author, but the practice of the editor," was Sansone's verdict,[78] which was in general terms a rehearsal of a criticism already voiced by Tore Janson as early as 1964.[79]

Sansone's judgement was met by a considerable number of additional critics, who now not only started to question the method, but the general truthfulness of the result that the *HA* was written by a single author. New arguments were brought to bear. Marriott would not have sufficiently trialed the general validity of his method by taking into account texts of a similar genre, which Meißner[80] and Frischer et al.[81] proved to in fact exhibit very similar length patterns, although written by different authors. Finally, Gurney and Gurney applied a battery of stylometric tests to the case study and showed that different stylometric tests would lead to different conclusions.[82] Joseph Rudman —a leading critic of all stylometric practice— concluded the series by raising with some indignation: "what are the 'non-technical' classicists to believe. The danger is that all of the non-traditional studies are now suspect."[83] Ultimately, in 2016, Stover and Kestemont had gathered enough courage to break the silence that ensued after Rudman's condemnation, and applied the latest innovative methods

[77] Ian Marriott, "The Authorship of the *Historia Augusta*: Two Computer Studies," *The Journal of Roman Studies* 69 (1979): 65–77.

[78] David Sansone, "The Computer and the *Historia Augusta*: A Note on Marriot," *The Journal of Roman Studies*, 1990, 174–7.

[79] Janson's criticism, found in Tore Janson, "The Problems of Measuring Sentence-Length in Classical Texts," *Studia Linguistica* 18 (1964): was especially voiced towards the publications of Louis Delatte and Étienne Evrard, who had begun work on indices to catalogue the concordances and frequencies of words in works by classical authors at the LASLA (Laboratoire d'Analyse Statistique des Langues Anciennes), notably with Seneca's *De Consolatione ad Polybium*. Delatte and Evrard's work was also based upon word length and lexical richness. See, for instance, Louis Delatte and Étienne Evrard, "Un laboratoire d'analyse statistique des langues anciennes à l'Université de Liège," *L'Antiquité Classique* 30–2 (1961): 429–44; and Louis Delatte et al., *Sénèque, Consolation à Polybe. Index verborum, relevés statistiques* (The Hague: Mouton, 1962).

[80] Burkhard Meißner, "Sum enim unus ex curiosis. Computerstudien zum Stil der *Scriptores Historiae Augustae*," *Rivista di cultura classica e medioevale* 34, no. 1 (1992): 47–79.

[81] See Bernard Frischer et al., "'Sentence' Length and Word-Type at 'Sentence' Beginning and End. Reliable Authorship Discriminators for Latin Prose? New Studies on the Authorship of the *Historia Augusta*," *Research in Humanities Computing* 5 (1996): 110–42; and Emily K. Tse, Fiona J. Tweedie, and Bernard Frischer, "Unravelling the Purple Thread: Function Word Variability and the *Scriptores Historiae Augustae*," *Literary and Linguistic Computing* 13, no. 3 (1998): 141–9.

[82] Penelope J. Gurney and Lyman W. Gurney, "Authorship Attribution of the *Scriptores Historiae Augustae*," *Literary and Linguistic Computing* 13, no. 3 (1998): 119–31.

[83] Rudman, "Non-Traditional Authorship Attribution Studies in the *Historia Augusta*," 152.

from the field.[84] A novelty was that they approached the authorship of the *HA* as an authorship *verification* problem instead of an authorship *attribution* problem (the *impostors* method, see 3.4.3, p. 117). Stover and Kestemont's conclusion was, as to be expected, careful, but they were able to show that there seemed to be at least a very conspicuous stylistic discrepancy between the *HA*'s primary lives and secondary lives, granting a remarkable meaning to the much-debated gap appearing in the corpus, which provided evidence to its not being a "happenstance of transmission."[85] Were two authors responsible for having written the *HA*?

The case of the *HA* presents a fine exemplary history for computational stylistics, *not* because it is a success story, but because it tells a history of a field coming to age, enclosed between two firing lines: philology on the one hand, statistics on the other. It also demonstrates perfectly how computational stylistics was still in its infancy when it was already attempting to prematurely answer some of the canonical questions, and this might not always have worked to the field's advantage. If some found the field's mistakes excusable as youthful exuberance, others condemned it as bluntly arrogant. Some might still find the matters that were pressing for the *HA* equally difficult or unsurpassably complex: the quality of selected features and the method of selecting them, the fundamental problems of stylistic corruption and intervention underlying the transmitted texts, the necessity of a background corpus or control texts to increase the reliability of the results, the problems inherent to the nature of Latin as a language (i.e. its being a highly synthetic and inflected language), its seeming robustness and school-preserved artificiality (already as early as the middle of the first century AD) that seemingly stands in contradiction with individualism and personal style, its lendability for a profuse number of methodological experiments, etc.

### 2.3.2 Databases and Ensuing State of the Art

The *HA* debacle did not stop computational stylistics from expanding a certain influence, the opposite was true: the bluntness of the method and its creative potential provoked interest motivated both from skepticism as genuine faith in its applicability. Furthermore, a practical reality for classicists (ancient Latin and Greek) was the availability of data facilitating computational research.

Classicists are generally regarded to have taken something of a head start when it comes to the systematic management of Latin data within electronic environments. Suchlike databases, which provided query tools that allowed its metadata and mark-up to be consulted, were inspired by Roberto Busa's *Index Thomisticus*, and from the 1960s–70s onwards emerged in various guises. An early example is the Laboratoire d'Analyse Statistique des Langues Anciennes (LASLA) in Liège, which started as early

[84] Justin A. Stover and Mike Kestemont, "The Authorship of the *Historia Augusta*: Two New Computational Studies," *Bulletin of the Institute of Classical Studies of the University of London*, 2016, 140–57.

[85] Ibid., 154.

as 1961, later on to be accompanied by such undertakings as Gregory Crane's *Perseus* project in 1987,[86] and Brepols' Cetedoc Library of Christian Latin Texts (CLCLT) in 1991.[87] Crane called the classicist tendency of classification and digital archiving "a data-intensive enterprise."[88] Scholars of classical languages were, in other words, pioneers and early practitioners of what is nowadays understood as corpus linguistics. As the field of digital approaches to the humanities grew larger, and technological advancements took flight from the early 1990s and 2000s onwards, scholars who applied computational stylistics for Latin were able to rely on these rich electronic collections.

Evidently, computational studies for Latin often betrayed an interest for determining a text's authorship, and a substantial number of successful applications exists. One could think of Forsyth et al.'s work on Cicero's alleged *Consolatio*, which had only been known from fragments until a manuscript and edition appeared in the sixteenth century.[89] The three scholars applied computational stylistics to function word frequencies in order to demonstrate that the *Consolatio*'s first editor, the humanist Carlo Sigonio (c. 1524–1584), was by all likelihood the author, and if not him, it certainly had not been Cicero. Tweedie et al. in their turn were able to prove on the basis of 134 most frequent items that *De doctrina Christiana*, Milton's alleged theological treatise interspersed with heterodoxies, consists of a quite variable style that exhibits divergent patterns of behaviour in comparison with Milton's other Latin works.[90] This hot potato for Miltonic scholarship was shown to resemble Milton only from afar, and rose new questions as to the text's provenance. Word of new computational methods to detect stylistic fingerprints trickled down further into the long-standing authenticity matters for the Latin Middle Ages. For example, the dispute of the authenticity of the *Epistolae duorum amantium*[91] was addressed statistically and/or computationally

[86] Gregory Crane and Elli Mylonas, "The Perseus Project: An Interactive Curriculum on Classical Greek Civilization," *Educational Technology* 28, no. 11 (1988): 25–32.

[87] As its contents widened beyond patristic and medieval texts, Cetedoc changed its name to Latin Library of Texts (LLT), see also this thesis's introduction on p. 18. On Cetedoc's history and reception, see the article by its director Paul Tombeur, "*Vox Latina*: Belgian Initiatives in Data Processing the Intellectual Language of Europe, A.D. 197–1965," *Computers and the Humanities* 12 (1978): 13–18; and Ron W. Crown, "Comparing the *Patrologia Latina* and the CETEDOC Library of Christian Latin Texts Databases from a User's Perspective," *Journal of Religious & Theological Information* 3, no. 1 (2000): 85–109.

[88] Gregory Crane, "Classics and the Computer: An End of the History," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth, Blackwell Companions to Literature and Culture (Oxford: Blackwell, 2004), 46.

[89] Richard S. Forsyth, David I. Holmes, and Emily K. Tse, "Cicero, Sigonio, and Burrows: Investigating the Authenticity of the *Consolatio*," *Literary and Linguistic Computing* 14, no. 3 (1999): 375–400.

[90] Tweedie, Holmes, and Corns, "The Provenance of *De Doctrina Christiana*."

[91] For the *Epistolae duorum amantium*, see further down this thesis on p. 246.

(in order of appearance) by Orlandi and Cupiccia,[92] von Moos,[93] Jan Ziolkowski[94] and Francesco Stella.[95] Through function word analyses, Kestemont et al. were able to show that two of Hildegard of Bingen's latest visions, the *Visio de sancto Martino* and *Visio ad Guibertum missa*, have all the characteristics of her last secretary, Guibert of Gembloux.[96] Eder provided computational confirmation of the hypothesis that the thirteenth-century chronicle *Chronica Polonorum* (i.e. *The Polish Chronicle*), transmitted under the name of some Gallus Anonymus, was in fact written by an author with a Venetian background, the Monk of Lido, for which —again— frequent function words provided a firm basis.[97] The duo Stover and Kestemont (with occasional guest appearances) seem to have had a most productive year in 2016, when they published 4 articles for authorship attribution of Latin texts: Julius Caesar,[98] the *Historia Augusta*[99] and Apuleius (twice)[100] passed the revue. Finally, in 2017, two articles for non-traditional authorship attribution for Latin appeared, one of which on Bernard of Clairvaux and his secretary Nicholas of Montiéramey,[101] which will be further treated and elaborated upon in this thesis, and the other on Pliny the Younger's letter to Trajan (53–117) on the Christians (and the emperor's response).[102]

## The *Quaestio Latinitatis*

What can be gleaned from a detailed reading of the listed examples above, is that the rapprochement between computational stylistics and Latin philology presents some

[92] Matilde Cupiccia, "Progressi nello studio del cursus: I metodi statistici e il caso di Eloisa e Abelardo," *Filologia mediolatina* 5 (1998): 37–48; Giovanni Orlandi, "Le statistiche sulle clausole della prosa. Problemi e proposte," Filologia mediolatina, 5 (1998): 1–36; Giovanni Orlandi, "Metrica e statistica linguistica come strumenti nel metodo attributivo," *Filologia mediolatina* 6-7 (1999): 9–31.

[93] See von Moos's counts of the different types of prose rhythms in the *Epistolae duorum amantium* on 43: "Anders als die Reimprosa zeigt der von Könsgen vernachlässigte Cursus-Gebrauch keinen Unterschied der beiden Rollen; beide bevorzugen eindeutig etwa gleichmäßig Planus (23%) und Tardus (21%) und vermeiden gemeinsam Trispondiacus (9%) und Velox (4%)."For the statistics, see the appendix on frequencies of the *cursus* at 103. His conclusions were that the two correspondents share their preferences in *cursus* usage, which is moreover completely opposite to the original letter correspondence of Abelard and Heloise. See Peter von Moos, "Die *Epistolae duorum amantium* und die säkulare Religion der Liebe. Methodenkritische Vorüberlegungen zu einem einmaligen Werk mittellateinischer Briefliteratur," *Studi medievali* 44, no. 1 (2003): 1–116.

[94] Jan M. Ziolkowski, "Lost and Not Yet Found: Heloise, Abelard, and the *Epistolae duorum amantium*," *The Journal of Medieval Latin* 14 (2004): 171–202.

[95] Francesco Stella, "Analisi informatiche del lessico e individuazione degli autori nelle *Epistolae duorum amantium* (XII secolo)" (Oxford, UK, September 2006), 1–10, `http://www.tdtc.unisi.it/digimed/files/EDA-Statistiche%20Lingua.pdf`.

[96] Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century."

[97] Maciej Eder, "In Search of the Author of the *Chronica Polonorum* Ascribed to Gallus Anonymus: A Stylometric Reconnaissance," *Acta Poloniae Historica* 112 (2015): 5–23.

[98] Kestemont et al., "Authenticating the Writings of Julius Caesar."

[99] Stover and Kestemont, "The Authorship of the *Historia Augusta*."

[100] Stover and Kestemont, "Reassessing the Apuleian Corpus"; Stover et al., "Computational Authorship Verification Method."

[101] De Gussem, "Bernard of Clairvaux and Nicholas of Montiéramey."

[102] Letters X.96–7, see Tuccinardi, "An Application of a Profile-Based Method."

interesting possibilities for cross-pollination and bi-directional advantages. This is no unique privilege to Latin per se, but one can hardly shake off the impression that Latin as a field and as a language has singular challenges to offer to computational methods which other Western-European languages cannot. Latin, both linguistically as historically, soon proved to be valuable for the improvement of the field of authorship attribution, as the language's linguistic peculiarities and long history insistently posed challenges unmatched by modern natural languages. Without a single exception, stylometric experimentation for Latin needed to engage with supplemental questions that are language-specific, or inherent to the textual culture in which the bulk of Latin texts was produced.

Forsyth et al.'s publication was one of the first "to extend the applicability of methods [...] to an *inflected* language, namely Latin," thereby emphasizing the novelty of the language being synthetic, and forming a challenge for Mosteller and Wallace's function word theory (2.2.2, p. 69): rather than using lexical items (function words) to designate the functions of words in sentences, Latin makes use of inflections and case endings. Other inherent obstacles for computational analysis of the language are of a historical-cultural kind, and one may well think back of the outlines given in chapter 1. Ziolkowski, in his discussion of the *Epistolae duorum amantium*, stressed that Latin was a language applied quite differently for different genres, and that genre-specificity could perhaps more easily than other languages take the upper hand when it comes to stylistic signals.[103] Such would have been a natural result of Latin's fossilization and school-taught terminological specialization. Stella's experiments on such generic specificities for epistolography,[104] as Eder's experiments on both ecclesiastic Latin[105] as Neo-Latin,[106] indeed suggest that the problem of Latin genre is a fundamental one. Forsyth et al. likewise questioned the model of imitation inherent to Latin's development as early as the beginning of Neo-Latin and Francesco Petrarch's (1304–74) advocation of a classical norm to re-establish the language's Golden Age: "At the heart of our problem, therefore, is the need to find a way of distinguishing between Cicero and Ciceronianism."[107] In this same regard, it should be mentioned that most Latin literature (be it ancient, medieval or Neo-Latin) has a prosodic dimension (be it different depending on the period).[108] In traditional scholarship as well, the

[103] "There are obvious shortcomings to the idea of comparing love letters with hymns, logical treatises, or theological writings," see Ziolkowski, "Lost and Not Yet Found," 191.

[104] Stella, "Generic Constants and Chronological Variations."

[105] Eder, "Taking Stylometry to the Limits."

[106] "It means that the hypothesis of chronological development of style has not been fully corroborated, even if some temporal patterns can be observed," see Eder, "A Bird's Eye View of Early Modern Latin," 71.

[107] Forsyth, Holmes, and Tse, "Cicero, Sigonio, and Burrows," 378.

[108] Classical Latin predominantly obeyed a quantitative rhythm, where the vowels within syllables were pronounced either short or long when scanned. The distinction between short and long also determined the linguistic meaning of words, which could play an important role as to the positioning within verse lines. From the late antique period onwards, Latin developed a more pronunciated stress accent, which trod in combination with the quantitative rhythm, possibly as a consequence of the weakening of the distinction between short and long syllables. This caused the increasing application of the medieval

*cursus* —the rhythmical closing of a sentence— has often been argued to appertain to stylistic preferences. Therefore, there have been some attempts to model Latin's musicality and rhythmic cadences as stylistic discriminants —by Linda Spinazzè, for instance—,[109] and apply them as stylistic features, most notably by Orlandi[110] and Matilde Cupiccia.[111] All of the above makes one wonder if Latin's obedience to norms and regulations presupposes a type of style that would transcend the personal level, one that is rather collective and shared.

Most Latin literature predates the industrialization of print and mass (re)production, wherefore most cases automatically come with transmission issues not commonly encountered in stylistic analysis of modern-day languages, or in forensic or sociological types of attribution studies. In such cases, the provenance and electronic availibility of social media messages, blogs and other types of web content, written in modern-day languages, often provide a much more secure basis. Latin provides a historical language, and this often means 'dirty data'. Examples of such concerns of contamination abound in the state of the art, or sometimes even provide the research question. Tweedie and others questioned whether or not the influence of several of Milton's amanuenses changing over time, such as Daniel Skinner and Jeremie Picard, might have caused the stylistic divergencies within his oeuvre.[112] Interestingly enough, they also emphasized that this given —the fact that Milton had stand-ins— caused the importance of their attribution to go beyond the mere act of attribution. Considering that the *De doctrina Christiana* had been used by several scholars as a gloss to interpret Milton's theological views, whether or not the text is truly attributable to Milton or only partly constitutes a vital element of the understanding of Milton as a historical figure in the context of seventeenth-century England. Equal concerns were formulated by Kestemont et al. in the attribution of Hildegard's suspect visions to Guibert of Gembloux: "it is clear that he reworked these texts so profoundly that hardly anything of Hildegard's writing style is still discernible in them."[113] Tuccinardi, in an earlier cited stylometric study contradicting the authenticity of Pliny's letter to Trajan regarding the Christians, had to make do with an edition from the early sixteenth century by Giovanni Giocondo (1433–1515) as an earliest witness. Although Giocondo appears to have relied on an ancient or medieval manuscript, the current unavailabil-

*cursus* —the rhythmical closing of a sentence— as applied by Latin-writing authors in prose, espcially from the second half of the tenth century to the first half of the twelfth century, see Tore Janson, *Prose Rhythm in Medieval Latin from the 9th to the 13th Century*, Acta Universitatis Stockholmiensis. Studia Latina Stockholmiensia 20 (Stockholm: Almquist & Wiksell International, 1975).

[109] Linda Spinazzè, "'Cursus in Clausula': an Online Analysis Tool of Latin Prose," in *Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem*, ed. Francesca Tomasi, Roberto Roselli Del Turco, and Anna Maria Tammaro, vol. 10 (New York, 2014).

[110] See Orlandi, "Le statistiche sulle clausole della prosa"; and Orlandi, "Metrica e statistica linguistica come strumenti nel metodo attributivo."

[111] Cupiccia, "Progressi nello studio del cursus."

[112] Tweedie, Holmes, and Corns, "The Provenance of *De Doctrina Christiana*," 83.

[113] Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 221.

ity of his source may make one question the reliability of Giocondo's transcription and edition in humanist times, when putting word of Christianity in the mouth of an authoritative, ancient author would have been most tempting indeed.[114]

## A Variety of Questions, a Variety of Markers

Bringing together the scholarly work surveyed above, one arrives at a long list of various markers or patterns by which attempts have been made to detect authorial style for Latin.

- Lexical richness: Yule's *K* measure was applied to *De imitatione Christi*.[115] Lexical richness was also applied in the early beginnings of the LASLA.[116]
- Sentence length: Both Yule as LASLA advocated for sentence length.[117] Marriott's attribution of the *Historia Augusta* brought the feature into some discredit.[118]
- Word length: Again by LASLA.[119]
- Function word *and/or* most frequent word (MFW) analysis: this has by far been the most popular method, although discussion has persisted as to whether or not it is useful to lemmatize[120] (refer the word encountered in a text to its dictionary headword) or not.[121] There is yet some different practice as to whether or not the top most frequent words have to be 'cleansed' from the content words, or if it is better practice to leave the process all-automated without any type of culling. The first stance is upheld by those who believe content words interfere drastically with the results,[122] whereas the second party believes that manual selection of the automatically generated top frequent words will yield undesirable external influence. The recommended minimum number of words fluctuates throughout the state of the art.
- Character / word / part-of-speech *n*-grams[123]

[114] Tuccinardi, "An Application of a Profile-Based Method," 436.

[115] Yule, *The Statistical Study of Literary Vocabulary*.

[116] Delatte and Evrard, "Un laboratoire d'analyse statistique"; Delatte et al., *Sénèque, Consolation à Polybe*.

[117] See Yule, "On Sentence-Length," and see n. 116 for the LASLA references.

[118] Marriott, "The Authorship of the *Historia Augusta*: Two Computer Studies," and see p. 72 where the *Historia Augusta* case was covered in full.

[119] See n. 116

[120] Gurney and Gurney, "Authorship Attribution of the *Scriptores Historiae Augustae*"; Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century"; De Gussem, "Bernard of Clairvaux and Nicholas of Montiéramey."

[121] "Preliminary tests for the Latin language have shown that lemmatisation, being a labour-intensive task, does not increase the attributive efficiency," see Eder, "In Search of the Author of the *Chronica Polonorum*," 12.

[122] Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 206–7.

[123] *N*-grams are sequences of *n* characters / words / parts-of-speech from a given sample of text or speech. Suppose that *n*=2, then a string such as "Sepe humanos affectus aut" could be respectively divided in character (`'Se'|'pe'|' h'|'um'|'an'|'os'|' a'|'ff'|'ec'|'tu'|'s '|'au'|'t '`), in word (`['Sepe humanos']|['affectus aut']`) or in part-of-speech tag (`['ADV NN']|['NN CON']`), where ADV stands for 'ad-

- Grammatically or syntactically annotated features: instead of analyzing word frequencies, some have suggested that observing the morphological and syntactic functions of words, and the order in which they appear, could be predictive of writing style. The problem is that such an approach requires more gunpower, as the text needs to be annotated. Annotation can bring along additional issues: it might introduce subjective judgement by the annotator, it can be very labor-intensive, or —if done automatically and the software is poor— it can be noisy and unreliable.[124]
- Word subsets: for instance, in the *impostors* method applied by Stover and Kestemont, 50% of the word frequencies were randomly selected in one hundred iterations.[125]
- Semantics: Zhang et al. applied a technique from distributional semantics. The full vocabulary of Caesar's war memoirs was projected into a lower-dimensional space by using random indexing (RI).[126]
- Prosodic features: respectively applied for the *cursus* (prose rhythm) by Orlandi and Cupiccia.[127] Verkerk and Turcan-Verkerk have likewise published on the statistical analysis of rhymed and rhythmic prose for the Middle Ages.[128] Forstall and Scheirer from the *Tesserae* project have statistically analyzed sound patterns in Latin elegiac couplets.[129]

Some technical background has thus far not been covered, wherefore the list can for now safely be interpreted as merely an illustration of the existent variety of selectable features. Equally so, different statistical techniques have lain at the basis of handling them, techniques which we have yet to discuss in the next chapter. What matters at this particular point is to illustrate the extant variety of applied methods. Inevitably, the question must rise which of these units is most reliable for segregating styles. If computational stylistics' aim is to present, in Moretti's words, "a more rational literary history,"[130] then what could be an objective guideline to determine which feature performs best and why?

Skeptics of stylometry have argued that the absence of explicit consensus within its practice compromises its validity. Such has evidently been the case for Rudman, who

verb,' `NN` for 'noun' and `CON` for 'conjunction.' *N*-grams are treated more comprehensively later on in this thesis, see the section on tokenization in A.3.2 on p. 313.

[124] Eder, "Taking Stylometry to the Limits."

[125] Stover and Kestemont, "The Authorship of the *Historia Augusta*," 146.

[126] Zhang, Cohen, and McGill, "Did Gaius Julius Caesar Write *De Bello Hispaniensi*?"

[127] Orlandi, "Le statistiche sulle clausole della prosa"; Orlandi, "Metrica e statistica linguistica come strumenti nel metodo attributivo"; Cupiccia, "Progressi nello studio del cursus."

[128] Philippe Verkerk and Anne-Marie Turcan-Verkerk, "Un Programme informatique pour l'étude de la prose rimée e rythmée," *Le Médiéviste et l'Ordinateur* 33 (1996): 41–8.

[129] Christopher Forstall and Walter Scheirer, *A Statistical Stylistic Study of Latin Elegiac Couplets*, November 2010, `https://www.researchgate.net/publication/251422159`.

[130] Moretti, "Graphs, Maps, Trees. Abstract Models for Literary History—1," 68.

has over the past twenty years criticized the disagreements between computational stylisticians.[131] Practically each of Rudman's articles concludes with a declaration of the field's invalidity. Since computational stylisticians cannot settle amongst themselves what constitutes good practice and what does not, it fails —in Rudman's eyes— in bringing across unambiguous resolutions to a non-initiated audience.

## 2.4 A Descendant of Textual Criticism

### 2.4.1 An Alternative History

The preceding subsection provided a history of computational stylistics according to the book, in which most of the field's beginnings are to be situated in the nineteenth-twentieth century, when the discipline's formalization or scientification took place. In this subsection I propose an alternative version. Computational stylisticians' empirical detection of mathematical principles underlying language, and Mosteller and Wallace's resolution to the phenomenon by drawing focus to small lexical items such as function words, was only revolutionary in a technical rather than in a literary-critical sense.[132] Despite a number of methodological innovations, the concerns for which the method is applied in many aspects repeats traditional-philological concerns towards the authenticity of historical sources. In other words: textual criticism as computational stylistics *avant la lettre*, or computational stylistics as a textual-critical *alter ego*.

Also in the field of textual criticism, the scrutiny of small, textual details, easily passed over in reading, would provide the textual critic with the key of uncovering the text's true nature. Computational stylistics can arguably be thought of as an intensification of that process. In a similar manner, the textual critic would trace and collect all minute evidence in order to grant a text its fixed reality and provenance (perhaps in the form of an edition). This similarity constitutes a problem, because computational stylistics thereby inherits the criticism that has weighed in heavily the past century on such traditional, philological models. Traditional philology was characterized by the reconstruction of an idealized, original text, one that would be placed in a hier-

[131] I give here a list of references to Rudman's articles, some of which have been cited before. See Joseph Rudman, "The State of Authorship Attribution Studies: Some Problems and Solutions," *Computers and the Humanities* 31 (1998): 351–65; "Non-Traditional Authorship Attribution Studies in the *Historia Augusta*," *Literary and Linguistic Computing* 13, no.3 (1998): 151–7; "Non-Traditional Authorship Attribution Studies: Ignis fatuus or Rosetta Stone?" *Bibliographical Society of Australia and New Zealand Bulletin* 24, no.3 (2000): 163–76; "The State of Non-Traditional Authorship Attribution Studies—2012: Some Problems and Solutions," *English Studies. A Journal of English Language and Literature* 93, no.3 (2012): 259–74; "Non-Traditional Authorship Attribution Studies of William Shakespeare's Canon: Some Caveats," *Journal of Early Modern Studies* 5 (2016): 307–28.

[132] Bod makes a similar point by asserting that different humanities disciplines, across periods and regions, appear to have had in common the will to discern principles and patterns in the textual material. See Rens Bod, *A New History of the Humanities. The Search for Principles and Patterns from Antiquity to Present* (Oxford, UK: Oxford University Press, 2013).

archy in which alternatives or variants were subordinated. Traditional philology was moreover restricted to description instead of interpretation. Once the stylistic peculiarities of each variant discovered and all elements identified, what yet remains to be done is a scholar's construction of an idea or a particular insight which amounts to more than describing what was already inherent to the data: *why* is there a difference? What does the phenomenon of difference imply for the literary culture(s) in which the text(s) were conceived? Description and differentiation of the sources —in the form of an edition or any other type of ascription— is only the first step. Understanding is the second. This constitutes a positivistic impasse, one which computational stylistics has inherited.

### 2.4.2 The Positivistic Impasse

Hugh Craig was concerned by seeing this problem and named an article after it: "If you can tell authors apart, have you learned anything about them?"[133] Although there is a very obvious philological merit in computational stylistics' main subtask, which continues to be authorship attribution, Craig's question implies that the results yielded by computational stylistics may often appear restricted to a matter-of-fact philological level, which envisions the recovery an original text (an archetype) by an original author.[134] In what follows I wish to further address the continuity between computational stylistics and traditional types of philology. I do this for two reasons. Firstly, this nuances the label of 'revolutionary' and 'innovative,' so often attached to computational methods in the humanities. Such assertions may be counterproductive and cannot always be taken seriously in consideration of the fact few aspects of the method are new from a conceptual point of view. As early as the fifteenth century Leon Battista Alberti (1404–1472) had already been counting vowels for Latin poetry in his *De componendis cyfris*.[135] The technological component of contemporary computational text analysis is undeniably impressive, but one may ask if this constitutes an achievement by information technology, not the humanities. The second reason why surveying the parallels between these two philologies proves interesting, is of a more optimistic nature: gaining better insight into finding this method's place in a long line of philological history might yield inspiring ideas as to where to direct the field's future trajectory.

[133] Craig, "Authorial Attribution and Computational Stylistics."

[134] Although it can be argued that there are directer computational descendants of the textual-critical method, such as digital stemmatology, digital collation and the automated production of critical apparatuses. See p. 10, n. 32.

[135] Bernard Ycart, "Alberti's Letter Counts," *Literary and Linguistic Computing* 29, no. 2 (2014): 255–65.

### 2.4.3 Scholasticism: Abelard

The tension of language and logic is patently the oldest and most pressing question of textual criticism, and computational stylistics inherited this burden. The struggle is noticeable already in one of the earliest forms of what we nowadays consider textual criticism, notably medieval scholasticism. Authorship attribution is of course much older and is no medieval invention,[136] but with the dialectics of the medieval schools discussed more fully in the previous chapter,[137] we discern some of the early attempts of formalizing and regulating the process of stylistic recognition and textual reconstruction. The shift in attitude takes place in the twelfth century, which —as Bynum noted— felt the need more than any preceding period in the early Middle Ages to identify, classify and evaluate: "A sense of models or types, a sense of proliferating groups and structures and of the necessity to choose among them, a sense of relationship, are characteristics of the twelfth century at least as salient as a new sense of self."[138] This is also reflected in the century's wide popularity and dissemination of *summa*, encyclopaedias, compendia and *specula*, which meant to summarize and classify sources in an increasingly systematic way.

Indeed, one of the earliest examples of the application of 'logic-driven classification' to textual criticism is Abelard's *Sic et Non*[139] —which, quite literally translates from Latin as something along the lines of `True` or `False`; or 'yes' and 'no.' Through a dialectic procedure reminiscent of proto-scholastic Anselm of Canterbury in his *Proslogion* (1033–1109) or the Bolognese jurist Gratian and his *Concordia discordantium canonum* (c. 1150), Abelard problematizes through the juxtaposition of 158 *quaestiones*, the existence of textual variants of Scripture and writings by the Church Fathers that seemingly stand in contradiction to each other. For instance, Abelard launches a *quaestio* (a theological assertion) such as the very first of *Sic et Non*, not coincidentally: "That faith cannot be proven by human reason, or on the contrary,"[140] and exhaustively lists sources (mainly Church Fathers) that both confirm or contradict the statement. These are not only diverse, but even adverse.[141] Frustratingly enough, Abelard provides no resolution to his own list. Only from his prologue do we gain an impression of what appears to have been his design, a design which he defends —among other instances— in this ensuing passage by allegedly quoting Jerome (if it were Jerome at all, the original is lost):

[136] Where exactly textual criticism originated, is of course difficult to tell, as there examples of textual-critical attitudes already abundant as early as antiquity. Harold Love, for instance, in the historical survey of his monograph on authorship attribution, situates the germination of text attribution and textual-critical attitudes with Athens in the fifth- to fourth-century BC, the Alexandrian editors and —for Latin specifically— Aulus Gellius, see Love, *Attributing Authorship*, 14–5.

[137] Efstathios Stamatatos and Moshe Koppel, "Plagiarism and Authorship Analysis: Introduction to the Special Issue," *Language Resources & Evaluation* 45, no. 1 (2011): 1.

[138] Bynum, "Did the Twelfth Century Discover the Individual?," 5.

[139] Petrus Abaelardus, *Sic et Non*, ed. Blanch Boyer and Richard McKeon (Chicago: University of Chicago Press, 1976–7).

[140] "Quod fides humanis rationibus non sit adstruenda et contra," in ibid., I. ll. 1–2, 113.

[141] "[...] non solum ab invicem diversa verum etiam invicem adversa videantur," ibid., *Prologus*, l. 2, 89.

> Similarly thus, "Let us plainly say how in Matthew and John it is written that the Lord was crucified in the sixth hour, whereas in Mark in the third. This was a scribal error, and in Mark the sixth hour was written, but many believe that instead of the Greek *episēmon* [F] there was a gamma [Γ]. Just so was there an error by scribes, that they wrote Isaiah instead of Asaph."[142]

In times of doubt, it takes a critical mind to determine —along the lines of applied logic— which of the variants has most authority. The logic in the preceding passage is that the capital letter of the Greek digamma (or episēmon), written as <F> and commonly used to denote the number 6, had been confused —by scribes— for the capital letter of gamma <Γ>, the third letter in the alphabet used to denote the number of 3. Such small details in the copying of the text can considerably alter the text's meaning and cause unwanted apocryphality. Spotting such anomalies in Scripture and criticizing them in the way that Abelard did was of course no sinecure. Abelard traverses between understanding the scriptural plurality as a deliberate and divine avoidance of *identitas* —stating everything in similar wording would be vulgar—, or as plain error. It is up to human *ratio* to tell which plurality is divine and eligible, and which plurality is human and corrupt. Textual criticism, in this prototypical, scholastic form, was born from a desire to detect what is deviant of an original, and a desire to detect which parts of the text had been contaminated through transmission, or even worse, were purposefully contorted or conflated to mislead its reader, or to lead away from divine truth.[143]

### 2.4.4 Humanism: Lorenzo Valla

In humanist times (fourteenth–sixteenth century) anxiety concerning concocted artefacts from the past reached new heights. Correspondingly, the textual-critical methods by which to detect these falsities were intensified. The impulse arose from the humanists' antiquarian concern for preserving authentic classical culture with an elegant Latin style that was harvested to serve and embellish contemporary cultural demands. Anything that did not fit this bill was more likely to be rejected as worthwhile or even authentic. The hypercorrective side effects of this humanistic attitude are, however, equally known, as well as the sophisticated forgeries of the time. Think, for instance, of humanist Carlo Sigonio's (c. 1524–1584) alleged forgery of Cicero's *Consolatio*,[144] or Desiderius Erasmus (1466/7/9–1536) and his forging of *De du-*

[142] "Item: 'Ergo simpliciter dicamus, quomodo scriptum est in Matthaeo et Iohanne quod Dominus hora sexta crucifixus sit, in Marco quod hora tertia, error scriptorum fuit, et in Marco hora sexta scriptum fuit; sed multi pro episomo graeco putaverunt esse gamma, sicut ibi error fuit scriptorum, ut pro Asaph Isaiam scriberent,'"in ibid., *Prologus*, ll. 67–71, 92. My translation.

[143] "Illud quoque diligenter attendi convenit ne, dum aliqua nobis ex dictis sanctorum obiciuntur tamquam sint opposita vel a veritate aliena, falsa tituli inscriptione vel scripturae ipsius corruptione fallamur. Pleraque enim apocrypha ex sanctorum nominibus, ut auctoritatem haberent, intitulata sunt; et nonnulla in ipsis etiam divinorum testamentorum scriptis scriptorum vitio corrupta sunt," ibid., *Prologus*, ll. 54–8, 91.

[144] Forsyth, Holmes, and Tse, "Cicero, Sigonio, and Burrows."

*plici martyrio*, which he presented as a patristic text written by Cyprian.[145] Humanist scholars' tendency to fill in the blanks as they occurred or to improve the mistakes of the past (very often the medieval past) occasionally meant an even worse blow to an authentic transmission of antiquity.[146] As both textual criticism's virtues and vices are represented from the beginning, the period provides a very clear antecedent to modern textual criticism and the detection of forgery, especially with such figures as Lorenzo Valla (1405/7–1457), who is often adduced as the direct ancestor to modern-day computational stylistics. In 1440, Lorenzo Valla published a treatise on the donation of Constantine (*Constitutum Constantini*), a document in which emperor Constantine (280–337 AD) would allegedly have transferred the authority over Rome and Western Europe to the pope. [147] Valla, on the other hand, declared the document —which had been accepted as genuine from the 9th to the fifteenth century— an early medieval forgery, probably written not long after the middle of the eight century.[148] His treatise, named the *Declamatio de falsa et ementita donatione Constantini*, drew attention to numerous anachronisms within the false document, and —more central to our current interest— small, lexical peculiarities and scribal inconsistencies.

As Coleman argued, "[...] for the first time, [Lorenzo Valla] used effectively the method of studying the usage of words in the variations of their meaning and application, and other devices of internal criticism which are the tools of historical criticism to-day."[149] There is an important nuance in Coleman's phrasing. Not so much Valla's recognition of the forgery, but the means by which he exposed it as such was what constituted a critical change. In a similar vein Alfred Hiatt argued that not the revelation of the donation as a forgery as such was what distinguished Lorenzo Valla from his predecessors, but "the mode by which texts are identified, criticized, and then classified as forgeries":[150] in other words, Valla developed a truly prototypical ancestor of textual criticism as we know it today. When computational stylisticians trace back their method's origins to Lorenzo Valla, —and indeed, many do— they have their ancestor in common with that of textual critics.

[145] Erasmus had fabricated the text to find support for his proper Christian ideologies. See Anthony Grafton, *Forgers and Critics: Creativity and Duplicity in Western Scholarship* (London: Collins & Brown, 1990), 44.

[146] "The new forgery stemmed less from practical needs than from nostalgia. It aimed above all at recreating a past even more to the taste of modern readers and scholars than was the real antiquity uncovered by technical scholarship. Many of the early recorders of monuments and inscriptions filled in missing texts in their notebooks just as they would the missing limbs and heads of statues, moved by an exuberant desire to see the ruined past made whole again; others, still less scrupulous, supplied whole new texts," see ibid., 26.

[147] After having been healed of leprosy by Pope Sylvester I, the emperor converts to Christianity, and grants the Roman See control over Rome, Italy and the western provinces.

[148] Although Nicholas of Cusa some 7 years earlier had "covered part of the same ground even better than Valla did, and anticipated some of his arguments," Christopher Bush Coleman, *The Treatise of Lorenzo Valla on the Donation of Constantine* (New Haven: Yale University Press, 1922), 3.

[149] Ibid.

[150] Alfred Hiatt, *The Making of Medieval Forgeries. False Documents in Fifteenth-century England* (London: The British Library, 2004), 16.

### 2.4.5 Mabillon and Lachmann

What we have here, of course, is computational stylistics' affiliation with a belief in principles of human rationality attached to the inheritance of humanism. Humanism is what provided the fertile soil for ensuing types of positivistic, philological study. Along with this conviction an arsenal of philological terminology arose, cleverly disguised in the natural-scientific shrouds of 'genealogical stemmas' or 'phylogenetics.' These were terms meaning "to obliterate the sin of scholarly subjectivity (*recensio, emendatio, stemma codicum*, critical apparatus)."[151] In Latin philology it offered prospects of imposing normativity and regulation to historical sources, that of classicism, which could draw its authority from scientific paradigms enabling to classify historical objects not into man-made categories but along 'verifiable' and 'objective' parameters. Valla kickstarted an approach that would shape the philology of the next centuries. Think of Jean Mabillon (1632–1707) and his *De re diplomatica* (1681),[152] the intellectual forefather of the critical analysis of medieval muniments. Mabillon was challenged by the Bollandist Daniel Papebroch (1628–1714) to defend the authenticity of Merovingian charters of the abbey of Saint-Denis. The undertaking required so much expertise that it resulted in his aforementioned classic handbook for Latin antiquarianism and contemporary diplomatics. Throughout six books, Mabillon devised a set of rules that would enable any careful scholar to distinguish the veracious from the false. A later and obvious example is Karl Lachmann (1793–1851), whose edition of Lucretius' *De rerum natura* laid the foundations for nineteenth-century German philology.[153] Lachmann's innovation was the principle that agreement in error implies identity of origin. In other words, he devised that a hierarchy of texts can be established according to origin, leading up to an archetype or urtext.

From Peter Abelard to Lorenzo Valla, from Jean Mabillon to Karl Lachmann, and ultimately to Frederick Mosteller and David L. Wallace: what has always been vital to these (proto-)textual critics was a sincere (or even sacred) concern for authenticity and truth as it is contained in our textual heritage, a bulk of texts of which we like to believe that they collectively provide passage to an accurate knowledge of historical ideas and events. For a textual critic, the final form of the text, how it is written, is never disjunct from its message or meaning,[154] wherefore 'form,' if we really want to seek out intention (not so easily detached from meaning at all), had to be 'original' and 'true.' These similarities announce the task of comparing computational stylistics' history of method to that of its ancestor. The question rises to what extent computational methods indeed present a 'new' philology, and where exactly the novelties are to be

[151] Richard Utz, "*Them* Philologists. Philological Practices and Their Discontents from Nietschze to Cerquiglini," in *The Year's Work In Medievalism*, vol. 26 (2012), 4–12.

[152] Mabillon, *De re diplomatica*.

[153] Titus Lucretius Carus, *De rerum natura*.

[154] One could argue that the very existence of a dichotomy of form and meaning is a priori false, and that since form and content are indissoluble it makes little sense to consider each of them separately at all.

located. If textual criticism can be regarded a forebear of computational stylistics — and I hope to have sufficiently demonstrated that such a case can be made— how does it also yield insights that textual criticism could not?

## 2.5 Conclusive Remarks

In 1974, quite early in the history of the computer, an article entitled "Computers and the Medievalist" was published in *Speculum*. For a moment, the article digresses into a discussion of the early experiments of computational stylistics on medieval texts (Latin and other languages). Amongst others, also Yule's work of 1944 (cited earlier) is discussed.[155]

> Without a clearly articulated theory, however, computational stylistics cannot fully develop. So far the theory has often been borrowed from quantitative linguistics [...] Quantitative linguistic concepts, in our opinion, are not easy to use; one should be very cautious with frequency counts and percentages, which are too often presented beyond any coherent theory of language.[156]

One might find it curious how little has changed. Some forty years later computational stylisticians are still struggling to embed their concepts theoretically and to find a way to approach the concept of style from what is there: frequency counts and percentages. Moreover, the methods by which such frequencies and percentages are attained lack standardized baselines and fixed protocols, and seem to be different dependent on the period, the researcher, the research subject, the language, the text, the author, etc. Clear articulation of what constitutes the method has remained a difficult point, and under way, a few blunders have discredited the field.[157]

[155] Yule, *The Statistical Study of Literary Vocabulary*.

[156] Vern L. Bullough, Serge Lusignan, and Thomas H. Ohlgren, "Report: Computers and the Medievalist," *Speculum* 49, no. 2 (1974): 397.

[157] In that respect, no method has been more notorious than Morton and Michaelson's *Qsum* method, a quite painful blemish to computational stylistics, as the method was applied in forensics (texts forged with criminal intentions and pursued in court). *Qsum* analysis is performed on the sentence level. The idea was that the authorship of a text can be detected in the correlation between sentence length and the frequency of two- and three-letter words. In other words, it takes the cumulative sums of the respective differences between sentence lengths in a sequence and the average sentence length (*qsld*), and correlates it with the cumulative sums of the respective differences between the number of 2-3-letter words per sentence in a sequence and the average number of 2-3-letter words across all sentences (*qs23w*), thereby embracing the idea that shorter words, which are often function words, are to be taken into regard as well. See Andrew Q. Morton and Sidney Michaelson, *The Qsum Plot*, Technical Report CSR-3-90 (Edinburgh: University of Edinburgh, 1990); Jillian M. Farringdon et al., *Analysing for Authorship: A Guide to the Cusum Technique* (Cardiff: University of Wales Press, 1996); The *Qsum* method was criticized extensively, most systematically and transparently in Stephen Karian, "Authors of the Mind: Some Notes on the QSum Attribution Theory," *Studies in Bibliography* 57 (2005): 263–86; Karian showed the method's lack of theoretical justification, how the authors empirically observed a correlation merely on the basis of a visual homogeneity and took no care —both statistically as qualitatively— to interpret the result. Karian's finishing stroke was to demonstrate that *Qsum* was based on a mathematical fallacy, and that *Qsum* led to different results for the same texts. The outrage and disgrace followed by the community was all the more exponential since the method had already been exploited in various court cases, whose verdicts had been decreed before consulting experts on the method's validity. See Juola, "Authorship Attribution," 243. In that sense, the incident was not only an embarrassment for those involved in

As long as computational stylistics is a field in development, it will be difficult to formulate theoretical tenets. Theories demand time, and over time they are refuted. It is important to emphasize that it is the humanities scholars who predominantly call for a theoretical framework, since this is part and parcel of what they believe to be sound scientific practice. Statisticians, on the other hand, work from a far more empirical and experimental approach, and are far more interested in the question if a model succeeds at measuring what needs to be measured. The tension and alienation between these two types of Weltanschauung is likely to stay, but, as Burrows said —some thirty years ago— when pondering upon the use and uselessness of a method still in the exploratory phase: "[...] The community would be the poorer if those who were not genuinely fluent in foreign languages were afraid to go abroad; if those who were not trained literary critics were persuaded that they should not read; and if those without thorough mathematical training were induced to avoid statistical analysis. It seems preferable to make the attempt, to take the obvious precautions, and to learn from the inevitable blunders."[158]

Even in the exact sciences there are various sides to the same story. There is no pressure to stick to one answer: statistics are not so much about ground truths than they are about probability. We can conclude this chapter by emphasizing how computational studies in the humanities are often driven by a means of finding *consensus*, a "consensus ex (or *qua*) machina."[159] *Consensus* is a very different concept than the establishment of an absolute truth. As Meister argued, it is "a strictly methodological consensus concerning the form of a literary text as opposed to a consensus on its meaning."[160] To have scholars consent upon verifiable and measurable principles, so as to found debate upon common observations, may prove to be computational stylistics' main epistemic contribution to the humanities in the future.

computational methods to language, but for science in general, and generated a good deal of suspicion as to how easily both academics and the non-initiated are prone to uncritical acceptance of objectified nonsense shrouded in an aura of scientific reliability.

[158] Burrows, *Computation Into Criticism*, 217.

[159] This was the title of the annual colloquium of ACH (Association for Computers and the Humanities) and ALLC (Association for Literary and Linguistic Computing) organized at the Sorbonne in 1994, see Marie-Anne Polo de Beaulieu, "Colloque *Consensus ex machina*?," *Histoire & Mesure* 9, nos. 1–2 (1994): 171–3. It was further explored in Jan Christoph Meister, "Consensus ex Machina? Consensus *qua* Machina!," *Literary and Linguistic Computing* 10, no. 4 (1995): 263–70.

[160] Meister, "Consensus ex Machina? Consensus *qua* Machina!," 263.

# 3

# Capturing a Stylistic Profile

## 3.1 Methodological Exploration

### 3.1.1 Outset

The purpose of this chapter is to go into confrontation with the methodology on a very practical level, and explore its strengths and weaknesses. We will be discussing how twelfth-century Latin texts are preprocessed, vectorized, sampled, visualized and analyzed. Throughout, the statistical encumbrances and difficulties will be addressed and problematized. In effect, as we discuss methodological questions and dig deeper into how to answer them, we will gradually progress from the more basic understanding of computational attribution to the more advanced. Then again, for an expert statistician this chapter might still be elementary, whereas the historian or literary scholar might feel that (s)he has to step up the game quite intensively. This will be a matter of background. As elsewhere in this thesis, I have nevertheless attempted to find a balance between eye for statistical detail and transparency, and the humanities scholar's disposal towards qualitative and language-specific interpretation of the results. In light of this I have chosen to append the most technical aspects at the back of this book to avoid all too much distraction (see p. 308 onwards).

### 3.1.2 The Benchmark Corpus

The intuitions of this chapter will be given step by step by drawing on hands-on examples from the benchmark corpus (tabularized in the thesis's appendix, see A.I.1, p. 304).[1] The benchmark corpus lays out an intricate network of the twelfth-century monastic and schooled writers in the region of contemporary France and Germany. An advantage is that through inductive experimentation on contemporary authors such as Rupert of Deutz (c. 1075–1130), Honorius Augustodunensis (c. 1080–c. 1154), Peter the Venerable (1092–1156), Hugh of Saint-Victor (c. 1096–1141),

[1] The division of corpora maintained in this thesis was introduced earlier on p. 18.

Guibert of Nogent (c. 1053–1125), and many others, we introduce some of the concepts and difficulties that are proper to computational analysis of twelfth-century Latin literature. Secondly, involving a wider corpus offers a hands-on introduction to the common misperceptions and problems of the stylometric method, as it also offers familiarity with the eleventh- and twelfth-century context in which our case studies are framed. In the same spirit of hardening the problem comes the additional aspect of shared and mutual influence that ties together the benchmark corpus. These abbots, teachers and saints operated in similar intellectual milieus, drew on similar sources, were similarly schooled, wrote on similar themes, and could occasionally stand into direct contact with each other, be it through teaching or correspondence. This means that common literary forms were cultivated, forms which could be altered and adapted toward the generic requirements of the text.

Another advantage is that some authors in this corpus left us oeuvres better suited to explore the difficulties of medieval Latin literature. Alan of Lille (c. 1128–c. 1203), for instance, is well-known to have explored most of the genres prevalent in his time, and deliberately experimented with them in subject matter as in form. The flexibility of which Alan of Lille and consorts testify considerably hardens the problem of unveiling a clean, purely authorial signal. All of this serves to emphasize that even though the benchmark corpus provides a frame of reference, which I will use through this thesis as illustrative footnotes to questions of authorship, it is not simply a gold standard. A 'gold standard,' in statistical terms, would imply that the confidence concerning authorship of all individual texts in the corpus is 100% guaranteed. Such a claim cannot be made, for historical reasons that have been more elaborately demonstrated in the foregoing chapters. I have nevertheless attempted to be as thorough as feasible concerning provenance, transmission and editing of the texts in this benchmark corpus (see column 3 of A.I.1, p. 304).

## 3.2 From Latin to Numbers

### 3.2.1 Preprocessing

In an initial phase, Latin texts are to be 'preprocessed' and/or 'normalized.'[2] Preprocessing entails basic operations in which the data —our Latin texts— are tidied up and transformed to adopt similar norms of formatting. This is because even the most minor editorial differences on the character level run the risk of being treated as stylistic differences by the computer (e.g. `'per,'` and `'Per'` would be treated as different words). In all of the experiments conducted in this thesis, the preprocessing was limited to some orthographical normalization of the type enlisted on p. 315. These are simple operations on orthographical principles which —as Latinists will know— can

[2] A more elaborate explanation of preprocessing and normalization can be found in the appendix in A.3.2 (p. 313).

fluctuate, such as <j>'s to <i>'s, <v>'s to <u>'s, <ae>'s to <e>'s. Other such tasks can be the removal of punctuation, of capital letters, but they can likewise pertain to the deletion of chapter titles and/or other forms of structural markup that might interfere with the text's form. There are more complex examples of preprocessing tasks imaginable. One such task is tokenization (see A.3.2, p. 313), which corresponds to having the computer learn how it should recognize and split units within a given text. Usually, the lower units of analysis are simply words, but they could just as well be *n*-grams.[3] More complex and 'cleverer' examples of preprocessing tasks are lemmatization, PoS-tagging and parsing, all of which are more elaborately illustrated on p. 316. Most of the preprocessing tasks in this thesis will be limited to rather simple, orthographical alignments. The current tools for taking into account more sophisticated annotation layers for Latin have thus far not yielded any considerable improvement in the field.[4]

There might be those who subject that orthographic and editorial information are to be considered valuable instead of negligible. In alternative settings of computational analyses or explorations of medieval texts, this stance can be readily defended. A possible instance of this is the recognition of a particular scribe or copyist, who might have maintained peculiar orthographic tics. The value of the variety of medieval sources was chiefly emphasized by the New Philology, already discussed in chapter 1, where the variation between texts is regarded to be a vital component of the cultural-historical understanding of the text.[5] Indeed, not every digital approach to medieval texts will normalize all different existent types of orthography. Digital editions seek to comprise the varieties within a manuscript transmission, wherefore these small variations are catalogued instead of normalized toward a single, 'ideal' state of the text. Indeed, the question of scribal transmission, and the use and (mis)use of normalization in this specific context, presents a topic in itself.

### 3.2.2 Sampling

It is customary in computational stylistics to slice the text into smaller segments or chunks, or so-called text samples. Sampling yields the advantage of "effectively [assessing] the internal stylistic coherence of works,"[6] as it also allows for a more fine-grained comparative analysis with segments from external works. As sampling implies the decrease of the number of words to what is arguably a minimum working length

[3] *N*-grams are sequences of *n* tokens from a given sample of text or speech. If *n* equals 2 or 3, one will respectively encounter the terms bigram and trigram. *N*-grams are explained in somewhat more detail in A.3.2 on p. 314.

[4] As was Eder's judgement after a series of large-scale experiments in Eder, "In Search of the Author of the *Chronica Polonorum*," 12; I demonstrate this point more fully in the appendix on p. 318.

[5] Cerquiglini and Nichols are the movement's figureheads, see Cerquiglini, *In Praise of the Variant*; and Nichols, "Introduction: Philology in a Manuscript Culture."

[6] Maciej Eder, Jan Rybicki, and Mike Kestemont, "Stylometry with R: A Package for Computational Text Analysis," *The R Journal* 8, no. 1 (2016): 111.

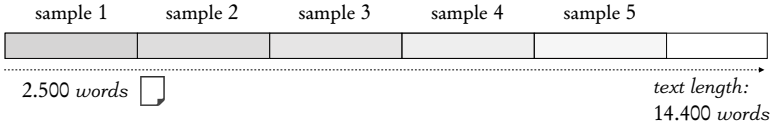



Figure 3.1: Intuition of sample length.

to still detect meaningful stylistic information, sample size has been a much-debated topic, also for Latin.[7] Deciding upon the authorship of a single sentence comes with a higher margin of error than for deciding upon the authorship of a paragraph, and deciding upon the authorship of a paragraph in its own turn is much more difficult than of an entire chapter, or an entire novel. In other words, the size of the population, and subsequently the size of the samples drawn from this population, are key in data analysis. The state of the art for computational stylistics in Latin has come to a consensus that once samples consist of some 2,500 to 3,000 words, the attribution accuracy of an experiment stagnates or reaches a point of saturation.[8] There are other types of sampling in computational stylistics, such as random sampling and generative sampling, more fully explained on pp. 318ff.

### 3.2.3 Vectorization and Feature Extraction

Once the data has become manipulable to the extent of the analyst's wishes, and all information is stored and categorized correctly, the final step of transformation is one in which 'tokens' are transferred to 'token frequencies.' This step is called vectorization, or is occasionally referred to as feature extraction. Gaining an intuition of this step from text to numerical values is often perceived as difficult, since it announces a stage at which the ensuing distant reading of the text will become somewhat more abstract: we will no longer be treating a text as if it consisted of words, but we will be analyzing a numerical representation of that text. Especially the statistical terminology intruding at this point is found to be tricky, although what is actually happening is quite straightforward. For the sake of clarity, fig. 3.2 yields an example in which the process of vectorization is approached intuitively.

Fig. 3.2 displays a hypothetical Latin corpus, consisting of three text samples, which each holds twelve tokens. Consequently, in the bottom of the figure, two rows appear. Row 1, entitled 'vocabulary,' sums up the entire vocabulary of the corpus (this

[7] Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 210–11.

[8] "Fig. 6 shows the result of this leave-one-out validation for various sample sizes (multiples of 100 lemmas, ranging from 500 to 4,000). It is obvious that larger sample sizes invariably lead to higher accuracies in cross-validation. Yet, whereas the initial accuracies are fairly low (even < 85%), the attribution success quickly rises above the psychological barrier of 95% (sample sizes > 1,500 lemmas) and becomes entirely flawless when dealing with sample sizes of ca. 3,000 lemmas or more." Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 214; Maciej Eder, "Does Size Matter? Authorship Attribution, Small Samples, Big Problem," *Digital Scholarship in the Humanities* 30, no. 2 (2013): 172.

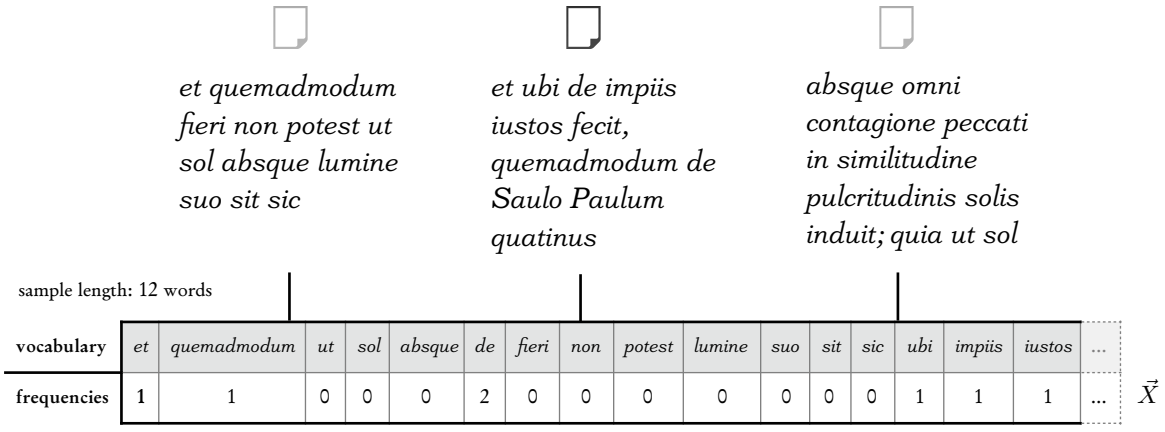



Figure 3.2: Intuition of bag-of-words (BOW) approach. Every document (or document sample) in the corpus is encoded as a vector $\vec{X}$, uniquely representing the frequency of each token within that document.

row includes the vocabulary of *all* samples, not just the middle sample). Think of row 1 as a tick list (or bucket list). Row 2 is the result of running over the bucket list with one specific sample in mind, namely the middle sample ("et ubi [...] Paulum quatinus"). The end goal is that for each unique text in the corpus the frequencies of the features are listed, which yields a unique new row —a vector representation $\vec{X}$— for each text sample. Note that the analyst can choose different ways of representing texts numerically. For instance, (s)he could have chosen not to take into account the entire corpus vocabulary, but only the 3 most occurring words in that vocabulary. As such, we can say that there are different feature extraction methods, and different possible derived values. This type of document representation is called a bag-of-words representation. Instead of respecting a document's original structure, words are aggregated and collected in a kind of 'bag,' disregarding any orderly principle of grammar or syntax.

The example in fig. 3.2 is obviously a dummy example, in which the statistical steps are easy to follow and can be recounted manually. When a text sample's length is only some 12 words, the statistical steps undertaken are easy to process for the human mind. Naturally, this process gains much more complexity when the corpus increases and contains more than 36 tokens. Simultaneously with text length, the corpus's total vocabulary from which to extract features will likewise increase, which in its own turn allows the exploration of additional 'dimensions' of texts in the corpus. The reader will be able to observe that the length of row 2 in fig. 3.2 is what will from now on be termed as the 'dimension' of the text representation. A two-word vector representation is called two-dimensional, because we can easily visualize it in 2 dimensions on a plot, such as in the left-hand side of fig. 3.3. Equally so, a vector space with a length of 3, will be three-dimensional (as can be gleaned from the right-hand 3D plot). As vectors get longer, they gain in dimensionality. As soon as this dimensionality rises above 3, humans have difficulties in visualizing the data patterns. Therefore, visualization in stylometry —and in data analysis universally— is still one

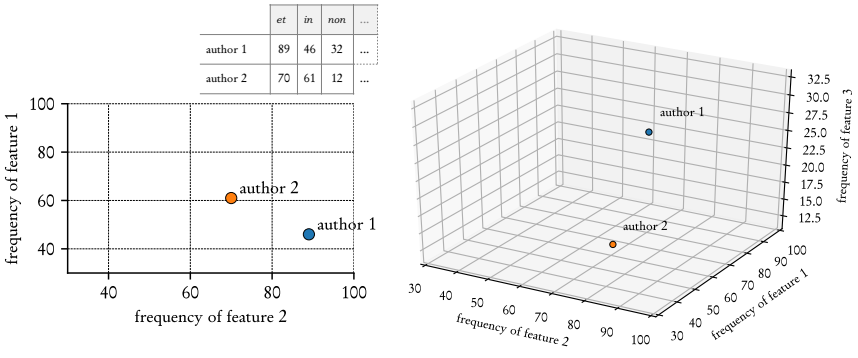



Figure 3.3: Intuition of vector space dimensionality.

of the field's main predicaments. We will return to this pickle in the field as soon as we come to discuss dimensionality reduction methods and visualization below on p. 96.

Once the text has been converted into numbers, the vectors are often scaled. This is because features which find themselves in larger and higher numeric ranges will easily weigh in heavier than those in lower ranges. In a sense, one can think of normalization as a necessity following up on Zipf's law (2.2.1 on p. 65) in corpora: top-scoring words reach very high peaks and dominate the statistics.[9] Scaling, then, is a means of forcing the values of all the feature variables within the same range. It sets a common ground upon which all observations are measured. Not doing it can have disastrous results. Here is not the place to discuss the various scaling methods. We will encounter a few variants throughout the thesis, and provide details where necessary. A more elaborate intuition of scaling is given in A.3.4 (p. 320).

### 3.2.4 Data Analysis

Once the text has become data, and is no longer human-readable, there are different kinds of statistical techniques by which to analyze the text's patterns. The following paragraphs will elaborate on two important concepts for analyzing data: classification and visualization.

#### Classification

Especially in the recent years, witnessing the rise of machine learning and computing power,[10] classification algorithms such as support vector machines (SVM's) have become increasingly popular. Classification is arguably at the heart of computational authorship attribution, i.e. assigning a label to a specimen of unknown origin, or categorizing according to class, "much like a spam filter nowadays is still trained on

[9] Piantadosi, "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions."

[10] Machine learning is introduced more elaborately in pp. 322ff.

| input vector $\vec{X}$ | | | | | | output label $Y$ |
|---|---|---|---|---|---|---|
| ['et' | 'in' | 'est', | ... | ] | | *author* |
| [-1.83 | -0.79 | -0.64 | ... | *n*] | — | 0 |
| [-0.38 | 1.19 | -0.25 | ... | *n*] | | 1 |
| [ 0.12 | 0.96 | -0.51 | ... | *n*] | | 2 |

Figure 3.4: Intuition of hypothetical data encoded in vector space (≈ text), which is consequently labelled according to class (≈ author).

manually annotated emails to learn how to distinguish between 'junk' email and normal messages."[11] The pioneers of computational authorship attribution, Frederick Mosteller and David L. Wallace, used a 'Bayesian' classifier in their landmark study on the Federalist Papers.[12] Given a set of classes (≈ known authors), we assign a class to objects that are unclassified (≈ anonymous texts). The gist of classification is that one first needs to feed a computer model with correctly labelled (preclassified) training examples. These training examples are computer-readable vector representations of texts (as demonstrated earlier in fig. 3.3). This type of learning, where the computer's predictions are founded upon 'seen' instances, is called supervised learning. In this scenario, "there is a set of variables that might be denoted as *inputs*, which are measured or preset. These have some influence on one or more outputs. For each example the goal is to use the inputs to predict the values of the outputs."[13] The process of this concept is explained in fig. 3.4. The style features' frequencies per author are fed (in fig. 3.4 these are 'et,' 'in,' 'est,' ...), their affiliation with an author learned, and consequently these same features can be reapplied as decisive discriminants for the attribution of a text of unknown authorship.

Classification is a huge field of research.[14] There are many types of classifiers, and it is not always clear which one will perform best, and why. Classifiers all react differently to different problems, have a variety of parametrization options and require other methods by which to optimize their performance during training. Therefore, they are typically evaluated as to their performance, so as to achieve a robust model which should avoid as much uncertainty as is statistically feasible.

It is generally easier to wrap one's head around classification by visualizing its actions from a geometric perspective. The visualization in fig. 3.5 allows to think of classification in spatial terms, where a document vector's values correspond to positional

[11] Stover and Kestemont, "The Authorship of the *Historia Augusta*," 144.

[12] Mosteller and Wallace, *Applied Bayesian*, ; for Naive Bayes, see p. 324. On the importance of Mosteller and Wallace for the field of stylometry, see p. 66.

[13] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2nd ed., Springer Series in Statistics (Stanford: Springer, 2009 (2001)).

[14] I go into more detail on different classifiers and their performances in the appendix to this book. See pp. 322ff.

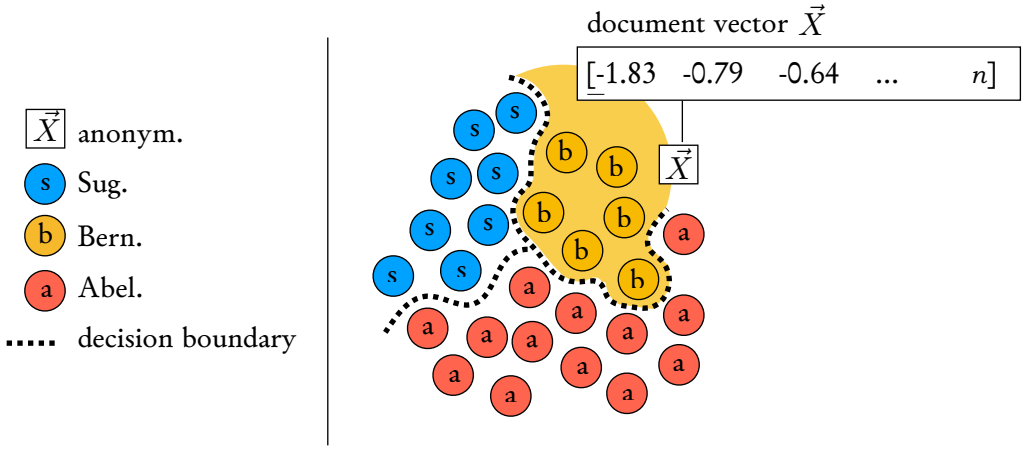



Figure 3.5: Intuition of decision boundary drawn by classification algorithm in between a hypothetical set of document vectors appertaining to Suger of Saint-Denis (blue), Bernard of Clairvaux (red) and Peter Abelard (yellow) to demarcate classes.

coordinates. Similar documents equal neighbouring data points in vector space, and classes are demarcated regions within that space. A document is categorized by the company it keeps, and by the space in which it is positioned. The frequencies of the features held within the document vector, then, allow a decision function to deduce from the array of values within a document vector how that document vector is most accurately classified. In the fig. 3.5, the anonymous snippet $\vec{X}$ is likely to have been written by Bernard of Clairvaux, because its coordinates make it fall on 'Bernard's side' of the decision boundary.

### Visualization

Visual inspection is essential to understanding data analysis, which the sheer number of figures in this thesis probably gave away. And yet, despite its potential to clarify and illustrate, visualization comes with deficits of its own. As was mentioned when discussing vectorization a few paragraphs earlier, data in computational stylistics often suffers from the curse of high dimensionality (as became clear from the number of columns in fig. 3.2). Humans, on the other hand, are only capable of grasping lower-dimensional (two- or three-dimensional) spaces. Hence, visualization techniques in stylometry, such as PCA scatter plots and networks,[15] need to accede to the limits of the senses. This immediately emphasizes why visualizations can be as helpful as they are deceiving: they are inherently always simplifications of much more complex realities. Founding attributive conclusions on the basis of visual inspection alone is

[15] See pp. 347ff. for a more detailed, technical explanation of PCA, and pp. 349ff. for Network Analysis. For the general use of PCA in computational stylistics, I refer to the classic José Nilo G. Binongo and M. Wilfrid A. Smith, "The Application of Principal Components Analysis to Stylometry," *Literary and Linguistic Computing* 14, no. 4 (1999): 446–66; a comprehensive introduction to network analysis can be found in Mark E. J. Newman, *Networks. An Introduction* (Oxford, UK: Oxford University Press, 2010).

precarious, to say the least. Most visualizations, or so-called clustering techniques, are moreover often unsupervised as opposed to supervised. The distinction is important. If we imagine 'supervised learning' as a scenario in which the computer is 'taught' by a teacher, unsupervised learning is a type of learning "without the help of a supervisor or teacher providing correct answers or degree-of-error for each observation."[16] The main goal of such techniques is description of a data set instead of strict classification, more of understanding patterns than of categorizing them. Therefore, it will often be more appropriate to perform multiple visualization techniques to back up an argument (consensus), or to back up a visual experiment with additional supervised evidence to corroborate the findings.

## 3.3 Trial Runs

Indeed, data analysis by means of a combination of supervised and unsupervised techniques is regarded sound practice. Such a transparent application of different techniques is largely a statistical answer to the problem of 'disagreement in stylometric practice' which formed the running thread throughout the previous chapter. Much of this thesis will indeed stand by establishing 'consensus' over different techniques, and remaining as open and transparent to the reader as feasible. An idea of why this is done, and what this looks like in practice, will now be demonstrated by concrete examples drawn from actual material collected in the benchmark corpus of medieval Latin texts at our disposal (A.I.1, p. 304).

Fig. 3.6 presents a series of PCA cluster plots where texts of several of our corpus authors are compared against each other. As is more fully explained on p. 348, PCA is an unsupervised clustering technique, which reduces the number of dimensions of the feature vectors to merely two or three principal dimensions of variance, or 'components' (PC's). Thereby, it only keeps the information that best reveals the most striking patterns in the data set.[17]

My selection of the particular text combinations in the subfigures was made solely on the criterion of which combination would yield the least class imbalance (where the number of words per author is roughly equal).[18] I have chosen to never combine more than three authors in the PCA plot, which is advised practice in the literature.[19] One can also observe that I have purposefully chosen a different set of parameters for each of these cluster plots, respectively alternating the *type* of features (character *n*-

[16] Hastie, Tibshirani, and Friedman, *Element*s, 486.

[17] For an elaborate explanation of PCA and its applicability to stylometry, see Binongo and Smith, "The Application of Principal Components Analysis to Stylometry."

[18] Class imbalance is a situation in which the proportionality of the classes is somewhat unequal, i.e. when the number of texts of one author far exceeds the other(s).

[19] "In stylometry, where indeterminate factors can influence variability, it is advisable to compare an anonymous text with only two authors' texts at a time," see Binongo and Smith, "The Application of Principal Components Analysis to Stylometry," 464.

grams, most-frequent words and restrictively function words), the *number* of features (from 100 to 1,000), the text sample length (from 500 tokens to 5,000) and the vectorization/scaling methods (standard-scaled raw frequency or tfidf-weighted frequency). This I have done to demonstrate out in the open that the use of different features to measure style affects the result. I can imagine that the reader might be interested in observing what happens when one alternates settings for one and the same combination of three authors, but we will return to this (p. 101). The explained variance below the plots summarizes how much of the data's original richness is explained by the two or three principal components. The more principal components are introduced to describe the data, the more the variance of the original data prior to the reduction can be 'explained' (this is why the percentage lies higher in fig. 3.6d, where the data is projected three-dimensionally).

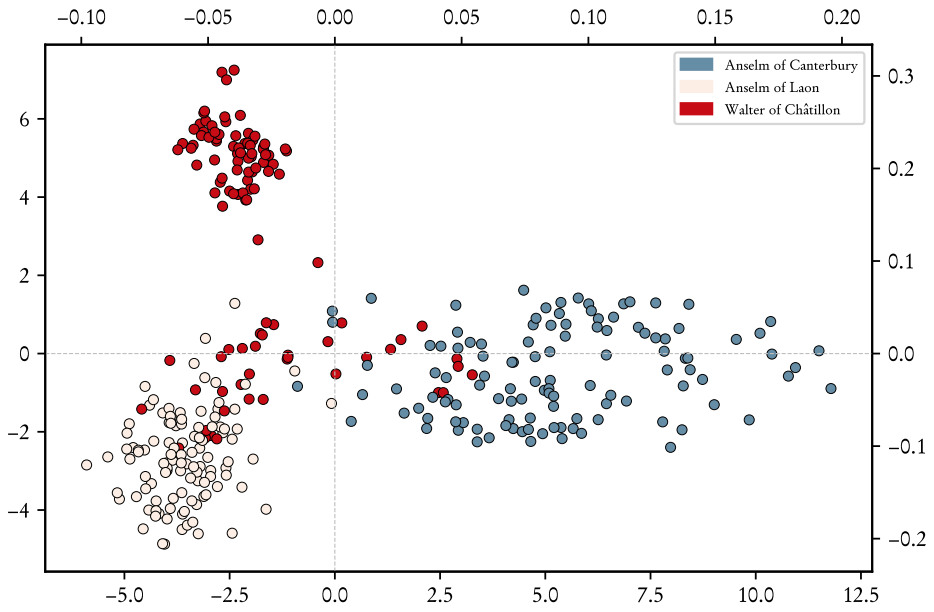
Let us address the simple question if computational stylistics performs for Latin at all, despite the many *caveats* inherent to the language and the historical context explored more exhaustively in the previous chapter. One can deduce that the answer to that question generally appears to be positive. Text samples of identical authorship (which, for the sake of convenience have been colorized in the scatter plots)[20] cluster together and seek each other's company in a two-dimensional or three-dimensional space. On the other hand, one can also notice many difficulties, which soon prove to be more interesting than those results confirming the method's performance.

For instance, Walter of Châtillon's epic poem, *Alexandreis*, which corresponds to the collection of red dots in the top of fig. 3.6a, wavers far off from his prose works. These two treatises, one polemic against the Jews and another on the Trinity, are so dissimilar from Walter's epic poem that they seek out the company of Anselm of Canterbury and Anselm of Laon's prose corpora. The two Anselms —despite their having been confused commonly in textual-critical history—, seem to be quite distinguishable for the computer. Walter of Châtillon's prose, however, dangles in between the two and cannot seem to find coherence nor proper place. Possibly, genre has played a decisive role here, as the second principal component neatly divides the figure in half: a lower half reserved for prose, and an isolated corner in the upper half reserved for Walter of Châtillon's poetry. Interesting in this respect is that Walter's shortest prose text, *De Trinitate*, was only very tentatively attributed to him by the Benedictine historian Bernhard Pez (1683–1735) in the eighteenth century.[21]

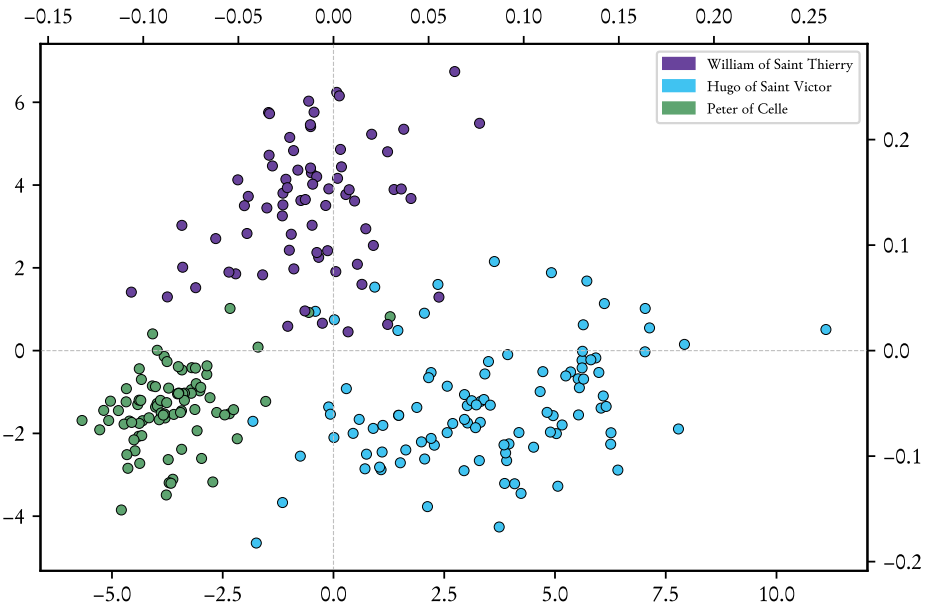
Fig. 3.6b demonstrates that authors such as William of Saint-Thierry, Hugh of Saint-Victor and Peter of Celle are distinguishable, but that Peter of Celle's style is more disciplined. Or, at least, it tells us this on the basis of the frequencies of the most common *function* words only. The corpora of William of Saint-Thierry and Hugh of Saint-Victor paint a quite different picture, one of incoherence. Many of the their

[20] Note that the algorithm is uninformed of the colour labels and is not in any sense manipulated.

[21] *De Trinitate* is part of an appendix to Walter of Châtillon's volume in the *PL*, and Pez's *monitum*, which was originally published in his 6-volume *Thesaurus anecdotorum novissimus*, can be consulted in *PL* 209:573–5.

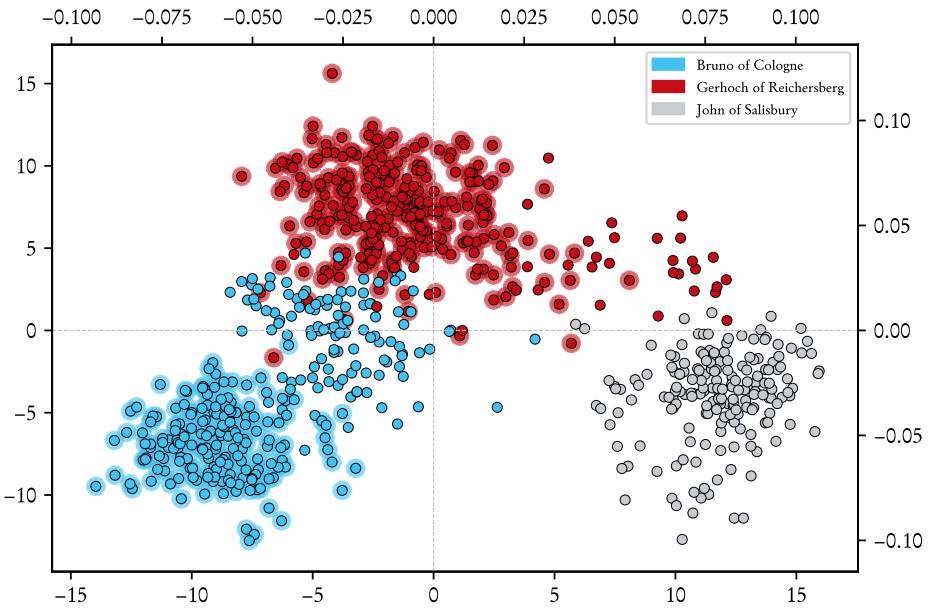



(a) *s–l* = 500 (!) | *type* = most-frequent word tokens | *n* = 200 | *vect.* = standard-scaled raw frequencies | *expl. var.* = 13.63%.

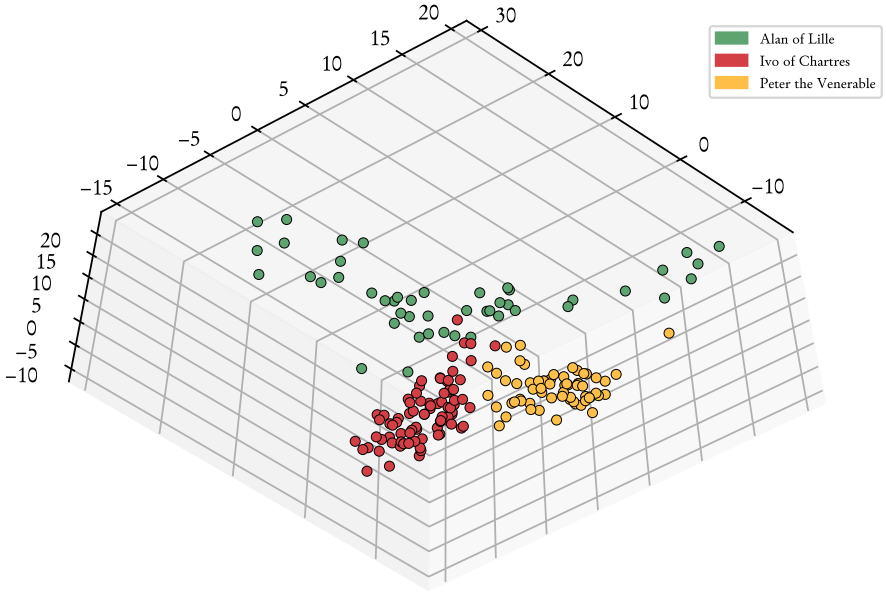



(b) *s–l* = 3,000 | *type* = most-frequent function word tokens | *n* = 100 | *vect.* = standard-scaled raw frequencies | *expl. var.* = 18.69%.

Figure 3.6: First two exploratory PCA cluster plots on eleventh- and twelfth-century texts from the benchmark corpus. Abbreviations: *s–l* = sample length | *type* = type of feature maintained | *n* = number of features | *vect.* = vectorization and scaling method | *expl. var.* = explained variance.




(c) *s–l* = 2,000 | *type* = most-frequent character trigrams | *n* = 1,000 | *vect.* = standard-scaled tfidf-weighted frequencies | *expl. var.* = 10.11%.




(d) *s–l* = 5,000 | *type* = most-frequent character 3- and 4-grams | *n* = 500 | *vect.* = standard-scaled frequencies | *expl. var.* = 29.82%.

Figure 3.6: Last two exploratory PCA cluster plots on eleventh- and twelfth-century texts from the benchmark corpus. Abbreviations: *s–l* = sample length | *type* = type of feature maintained | *n* = number of features | *vect.* = vectorization and scaling method | *expl. var.* = explained variance.

text samples are outliers or overlap with samples by other candidates. A similar trend of wandering text samples surfaces in fig. 3.6c, where strictly the most-frequent 3-grams were used for the corpora of Bruno the Carthusian (or Cologne), Gerhoch of Reichersberg and John of Salisbury. The authors' clusters are somewhat more dense, but near the edges, confusion between their respective stylistic signals is noticeable. What is interesting here, is that both Bruno and Gerhoch have written a commentary on the Psalms. In the figure, the data points of the text samples belonging to these commentaries are highlighted. In contrast to what one would intuitively expect, it is —in this particular case— not the common vocabulary of both Gerhoch's and Bruno's commentaries that has caused the overlap between their two respective clusters. The psalm commentaries are stylistically distinctive, and their similar topic does not cause the confusion near the clusters' edges. Finally, in fig. 3.6d, testimony is given of how difficult it is to drive Alan of Lille's stylistic signal into a corner. In this 3D-projection of the data's 3 principal components, which can be inspected from various angles, only a few perspectives allow to find a plane that distinguishes between the classes of these three authors. Throughout, Alan of Lille's versatile style is apparent, spreading over a wider space than Peter the Venerable and Ivo of Chartres combined.

### 3.3.1 The Multiplicity of Result

#### Four PCA Plots

The preceding prospection might strike one as quite an unusual way to be speaking about twelfth-century authors. The texts are reduced to merely visual impressions of their contents. The authors' thoughts, their complex backgrounds, their texts' transmission histories: numerous details considered important in a serious philological approach to authorship were abandoned in favour of larger patterns. The question now rises if these patterns allow for attribution success, and dependent on that question, what can be gleaned from them.

At first sight, there is some reason to foster optimism. In most scenarios, different features and sample lengths 'work' to a lesser or greater extent to detect authorial signals. Yet, the variety of ways by which to explore the texts' data and the potential links between text samples leaves open the question which set of parameters would be most advisable in an authorship attribution problem. Fig. 3.7 demonstrates this problem more fully.

In this figure, four times the exact same corpus consisting of texts by Guibert of Nogent, Peter of Celle and William of Saint-Thierry is revisualized with different settings. Although each of these visualizations appear successful to an extent in segregating these authors' texts, each of them is slightly different, and accentuates other aspects of *how* these authors distinguish themselves. This can be further emphasized by considering the following: let us pretend for a moment that in each of the four subplots a partition of the data had been withheld in advance, in order for it to function as test

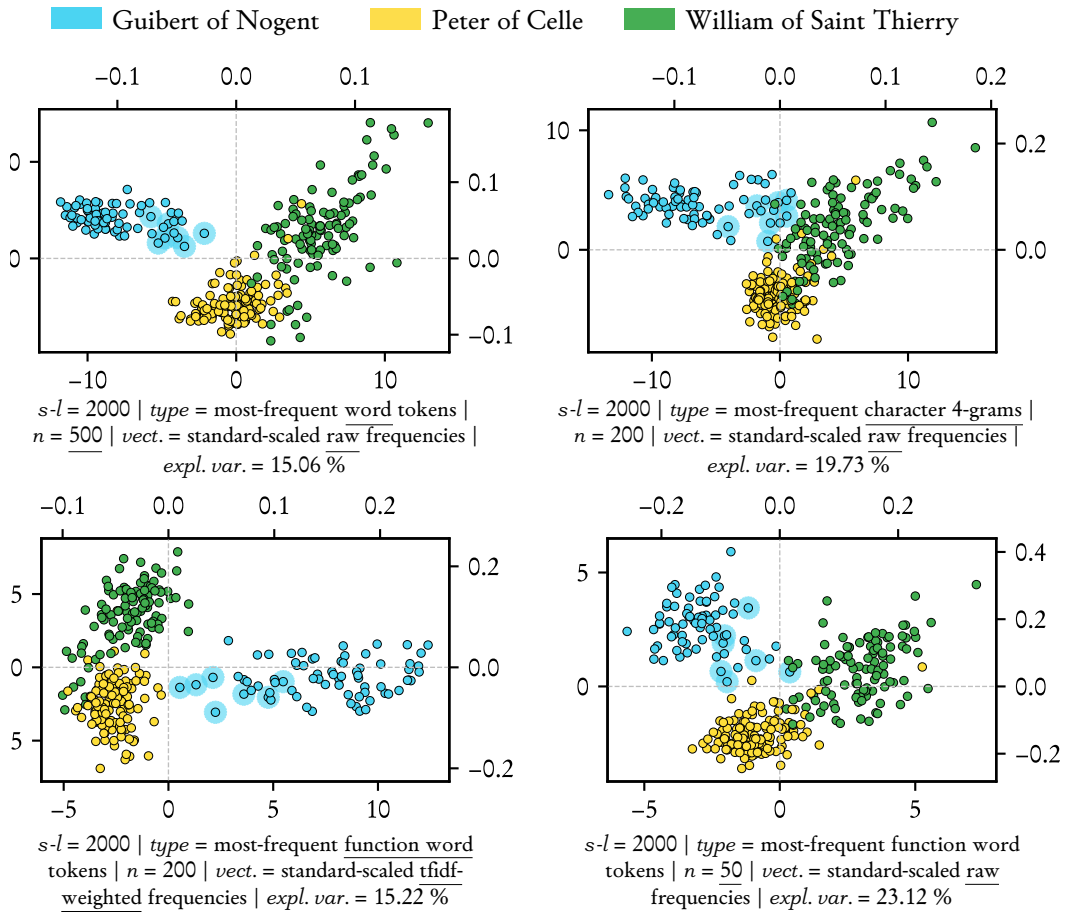



Figure 3.7: 4 PCA scatter plots of the same corpus generated with different settings. For the sake of comparability, a sample length of 2,000 tokens was maintained throughout. Abbreviations: *s-l* = sample length | *type* = type of feature maintained | *n* = number of features | *vect.* = vectorization and scaling method | *expl. var.* = explained variance.

material (a hypothetical specimen of unknown authorship). In fig. 3.7, these so-called 'anonymous' text samples —of which we know that they actually form part of Guibert of Nogent's *Contra iudaizantem et Iudaeos*— were highlighted (light blue). It is clear that the treatise, more than any other text by Guibert, shows some sympathies with the clusters of William of Saint-Thierry and Peter of Celle. But the degree by which this stylistic affinity is expressed somewhat differs dependent on the numerical representation one chooses to follow.

From an extreme point of view one could argue that these subplots lead to opposite conclusions when it comes to authorship. The question, then, is how to handle what one may come to denominate as a 'grey' zone. Clearly there is no prescribed or ideal setting by which to statistically determine authorship. As has been mentioned a number of times already, I will argue in the next few paragraphs by aid of more sophisticated techniques that the best solution consists in establishing consensus. 'Relevant' information can be excluded from the 'redundant' by performing an experiment over and over again, under different circumstances and with different settings. In the end,

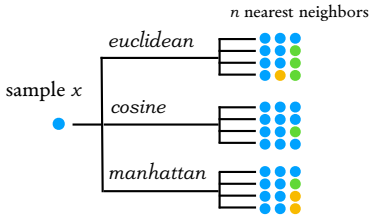



Figure 3.8: Intuition of a consensus network. For each sample, different parameters are tested. Each of the combinations yields a top three (the *n* nearest neighbours), which can consequently be averaged.

it is common sense to lend ear to a majority.

### 3.3.2 A Twelfth-Century Consensus Network

Indeed, when looking at the four scatter plots in fig. 3.7, the idea of 'preferring a majority' in order to achieve a consensus can already be brought into practice. Only the top right figure truly casts doubt over the authorship of *Contra iudaizantem*, wherefore three quarters of the time there is no reason to deny Guibert the credit of having written the text. The gist of achieving consensus and minimizing doubt indeed boils down to this exercise, be it in a more intensified manner.

What would 'building up a consensus' and 'minimizing doubt' look like in practice? Certainly one could decide to generate many different scatter plots with different settings such as in fig. 3.7 in order to gain a glimpse of the bigger picture. After all, none of these figures is strictly true or false, but each one of them provides a different representation of the same phenomenon. However, this is time-consuming and computationally expensive. In this light, the analyst might find it convenient to combine and integrate many different snapshots into one powerful visualization. Maciej Eder has recently experimented with the limitations of visualization in stylometry, and proposed network analysis[22] as a good alternative for gaining an overview of the stylistic affinities between a large number of texts.[23] Since the edges of a network are essentially containers which can be charged with any numerical value, Eder devised the concept of a consensus network, which is schematically depicted in fig. 3.8.

The idea is that a text sample's 3 'nearest neighbours' are calculated by iterating over different parameter settings. In a first phase, the algorithm computes the closest neighbours on a superficial level by looking at the 50 most common words. Consequently it increases this number by calculating that distance for 150, 500 and ultimately even 1,000 most common features, which captures textual connections that are no longer superficial but reach far down the frequency strata where thematically or generically

[22] See p. 349 for a detailed explanation of network analysis for stylometry. For a comprehensive introduction to network analysis, I refer again to Newman, *Networks. An Introduction*.

[23] Maciej Eder, "Visualization in Stylometry: Cluster Analysis Using Networks," *Digital Scholarship in the Humanities* 32, no. 1 (2017): 50–64.

linked word variables contribute to the overall picture. Additionally, one could also calculate the 3 nearest neighbours through the different distance measures introduced in A.3.5 (p. 333): euclidean, cosine and manhattan. The result is that a single text sample will end up 12 times (3 distance measures x 4 feature vectors with increasing dimensions) with top-3-candidates of nearest neighbours (see fig. 3.8). From this pool of candidates, one can simply deduce an average, stable top 3 that is most consistent across the different parameters, thereby excluding any kind of relationship between nodes that is merely coincidental.

The result is a mosaic-like patchwork (see fig. 3.9) in which all the texts in the benchmark corpus are represented. Interestingly, one can see that the authorial signal remains predominant, as clusters generally group according to colour (the legend of colours on the left indicates the authorship of the text samples). On the other hand, thematic or generic links cause the appearance of alternative groupings of samples, where the authorial signal is largely smothered. In the bottom right corner, for instance, Bruno of Cologne, Gerhoch of Reichersberg and Peter Lombard's writings seek out each other's company because of their common subject of writing: the Psalms and the letters of Paul. The many references to these biblical source texts, and the constant alluding to and citing of passages that is typical to the exegetical genre, seem to have forged very strong ties between these authors' text samples.

Other writers' contributions, on the other hand, lie scattered over the network: William of Saint-Thierry's (dark green) and Alan of Lille's (light green) texts are too diverse to be identified. In general, the network can be approached roughly as representing a core and a periphery. Within the core, stylistic confusion reigns supreme, and authorial identity is anonymized in a colourful mass. The center of the network, nevertheless, constitutes the beating heart of twelfth-century religious literature as we know it. Especially the more active and influential writers of the twelfth century are represented. They are Peter of Celle, John of Salisbury, Peter the Venerable: figures with a wide network and correspondence. On the left-hand side, Ivo of Chartres' and John of Salisbury's texts show close resemblance, which is fascinating since both authors could never have met each other (John was born ten years after Ivo's death c. 1115). Nonetheless, their profiles are comparable. Both have a history connected to Chartres where they would become bishop[24] and have links to the intellectual milieu in and around the Norman abbey of Bec.[25]

[24] Ivo was bishop of Chartres from 1090 to 1115. John of Salisbury is thought to have been taught three years of grammar at the school of Chartres by William of Conches. He left around 1141. His time in Chartres fell in between two periods of studying in Paris, first under Peter Abelard and consequently under Gilbert of Poitiers. Much later (in 1176) John became bishop of Chartres until the end of his life. See Christopher N. L. Brooke, "Introduction," *The Letters of John of Salisbury*, ed. William J. Millor, Harold Edgeworth Butler, and Christopher N. L. Brooke, trans. Christopher N. L. Brooke and Harold Edgeworth Butler, vol. 1 (Oxford, UK: Clarendon Press, 2003 (1955)), ix–lxvii, especially from xii onward.

[25] Ivo studied in Bec under the influential abbot Lanfranc of Pavia (1005–89), in all likelihood with Anselm of Canterbury (1033–1109) as a fellow class member. Anselm later became abbot of Bec. See Bruce C. Brasington, "Lessons of Love: Bishop Ivo of Chartres as Teacher," in *Teaching and Learning in Northern Europe: 1000–1200* (Turnhout: Brepols,

The peripheries, on the other hand, are far more stable, and occasionally present us with the 'outliers' of the twelfth-century: nodes that show the least connections with the corpus are pushed to the outer bounds of the network. Interestingly, in some cases, they somewhat reinforce the commonly maintained impression of these writers' positions within the twelfth-century intellectual milieu. The Benedictine monk and abbot, Rupert of Deutz, was the most prolific writer of the twelfth century, but remained somewhat remote (also geographically) from the early twelfth-century centers of learning. His traditionalist formation had been limited to that of his cloister, St. Lawrence (Liège),[26] where he probably would have stayed had accusations of heresy (impanation) provoked by certain passages in his *Liber de diuinis officiis* not forced him otherwise.[27] Generally, Rupert's writings were known, though to a considerably lesser extent than those of —for instance— William of Saint-Thierry or Hugh of Saint-Victor. His writings form a big, detached cloud in the upper right corner (brown). Other similarly isolated (and enigmatic) figures are to be found at the outer regions. That Honorius Augustodunensis is one of them (in orange) is hardly surprising, considering the effort the latter undertook to keep his identity and whereabouts uncertain. The place to which the epithet 'Augustodunensis' refers has in itself been a subject on which much ink has been spilled.[28] In the bottom left corner (in light blue) we find Guibert of Nogent, whose writings were close to unknown in his own day. Guibert's existence would probably have been a footnote of the twelfth century if not for the interest of later scholars towards his curious autobiographical text —*De vita sua*— and his historical account of the first crusade (1095–1099) —*Dei gesta per*

2006), 131. John of Salisbury did not study at Bec as such, but at Paris and Chartres (as indicated in n. 24). Nevertheless, during his life he stood in close contact with leading members of the Norman abbey and its network (John of Salisbury is indeed known to have had an expansive network of connections). John also wrote a *Vita* on Anselm of Canterbury, who is well-known to have been abbot of Bec until his election as archbishop of Canterbury in 1089. It was under Anselm that the abbey of Bec reached the peak of its intellectual allure. John also became secretary to Theobald of Bec (c. 1090–1161), a successor of Anselm. Theobald had likewise first been elected abbot of Bec in 1137 before becoming archbishop of Canterbury a year later. John became his secretary at the latest in 1148.

[26] Rupert's epithet signifies his abbacy in Deutz, which took place only in the final 8 years and one-half of his life. See John Van Engen, *Rupert of Deutz*, Publications of the UCLA Center for Medieval and Renaissance Studies 18 (Berkeley, CA: University of California Press, 1983), 12.

[27] In this regard, John Van Engen placed Rupert's conservatism and relatively 'closed' career in contrast to that of William of Saint-Thierry. See John Van Engen, "Rupert of Deutz and William of Saint-Thierry," *Revue Bénédictine* 93, nos. 3–4 (1983): 328.

[28] Despite the success of his prolific writings and their wide dissemination during his own lifetime, we know very little of Honorius Augustodunensis's life. He went through painstaking efforts to keep his name and whereabouts uncertain, possibly out of fear of being charged for heresy. Honorius most likely came from Regensburg, although he might have spent some time in Augsburg in Southern Germany. Where his name, then, derives from —be it Regensburg or Augsburg— has remained somewhat difficult to ascertain. The manuscript transmission strongly supports the thesis that Honorius came from Germany, and not from France (Autun), as was formerly believed. Especially Valerie Flint has written extensively on Honorius's career, legacy and impact in a series of articles in the *Revue bénédictine*. She has maintained that Augustodunensis must refer to Augsburg. See Valerie I.J. Flint, "The Career of Honorius Augustodunensis: Some Fresh Evidence," *Revue bénédictine* 82, no. 1–2 (1972): 63–86; "The Chronology of the Works of Honorius Augustodunensis," *Revue bénédictine* 82, no. 3–4 (1972): 215–42; "The Place and Purpose of the Works of Honorius Augustodunensis," *Revue bénédictine* 87, no. 1–2 (1977): 97–127; "Heinricus of Augsburg and Honorius Augustodunensis: Are They the Same Person?" *Revue bénédictine* 92, no. 1–2 (1982): 148–58.

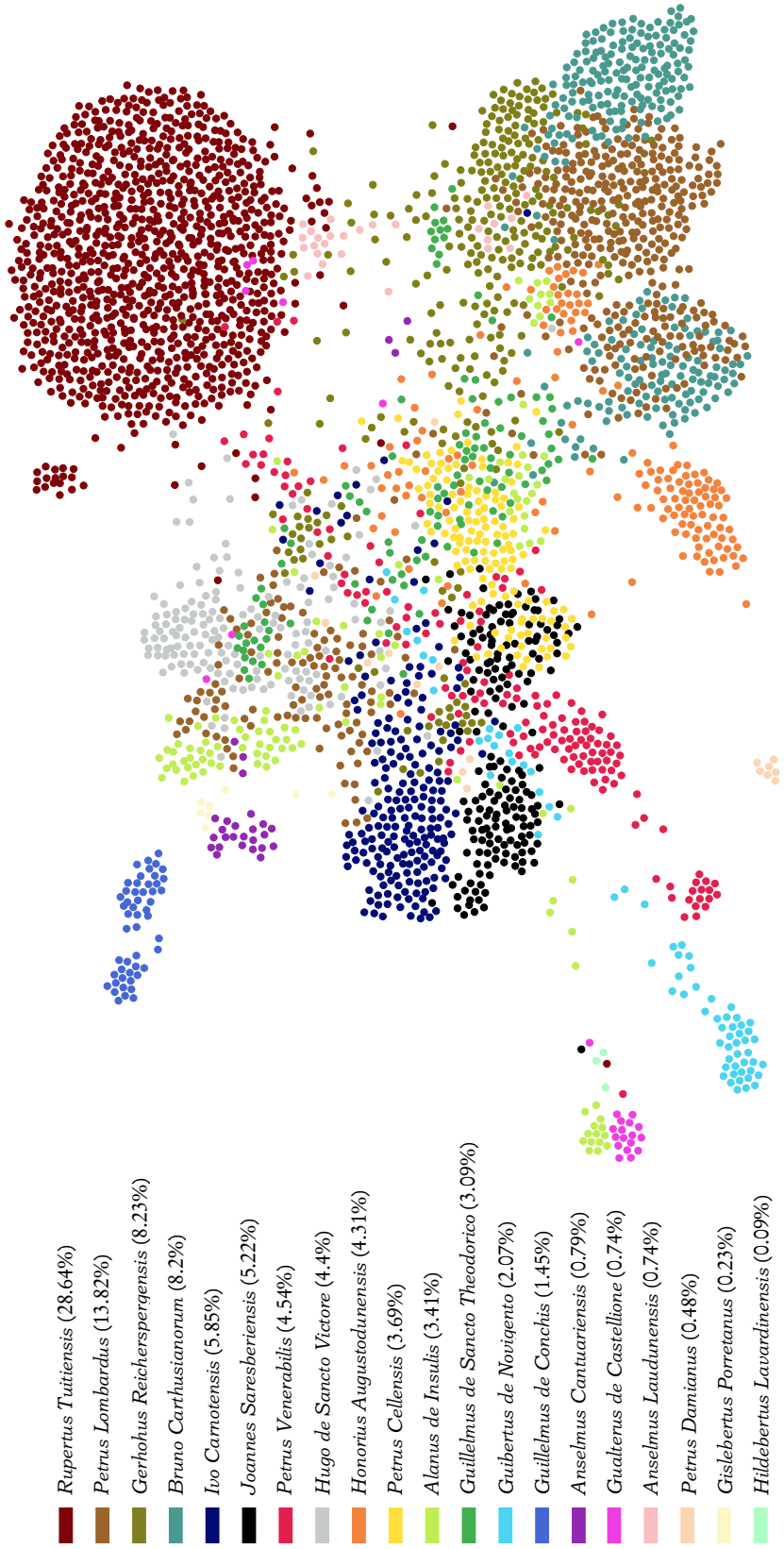



Figure 3.9: Consensus network of the benchmark corpus (see table A.I.1, p. 304). Sample length is 2,000 words. The connections between nodes were established for an increasing number of features (50, 150, 500, and 1,000 features). Three different distance measures were applied: euclidean, manhattan and cosine. The 3 nearest neighbours were computed on the basis of a naive brute force algorithm. Consequently, the nodes were connected with their neighbours and mapped in space through force-directed graph drawing (Force Atlas 2). For the sake of clarity, the edges between the nodes were not visualized, and overlap between the nodes was discouraged. The percentage next to the author name indicates its volume as opposed to the entire benchmark corpus.

*Francos*. Another outsider is William of Conches (dark blue, left top corner), one of the older *magistri* at the school of Chartres of whom very little is known, except that he exercised a great amount of influence on the disciples of the school of Chartres. Peter Damian (c. 1007–c.1073), who is often cited as a vital figure and inspiration for understanding the stirrings leading up to twelfth-century reform, interestingly enough shows very little connection to the writers of the next century (in apricot, at the very bottom).

### 3.3.3 Machine Learning: Order in the Chaos

Closer scrutiny of a consensus network is as instructive as it is puzzling. Its main asset is that it allows for comparison of a very large bulk of texts within a feasible scope, which can provide some general intuitions of the corpus and potential links between the candidates. Simultaneously, the fact that it simply combines *a lot* of textual information that can be statistically harvested, causes it to exceed an exclusively authorship-related visualization. A consensus network generated as the one above, interesting as it may be, cannot exclude irrelevant textual material related to extra-authorial factors, nor can it allow an exploration of how the reliability of its settings stands or falters. The latter constitutes a general disadvantage of unsupervised methods.

What seems desirable, then, is an evaluation of which parameters work and why. At which point does sample length become crucially deficient for authorial signals, for instance? Which are the optimal features for segregating classes of different authorship (feature selection)? Especially cross-validation (A.3.5, p. 340) and machine-learning methods (A.3.5, p. 322) which have introduced measures of reliability have proven particularly fit to provide some consensus in these kinds of scenarios. The idea is that each new set of data provides its own unique patterns and challenges, which the computer needs to learn in a training phase. In a nutshell, the advantage of training a classifier is that it can be divided up in authentic data in a training set and a test set. Consequently, we can continuously test the algorithm after it has learned different settings, and see which one of these performs best.[29] This successive search-and-evaluate process enables the classifier to obtain an optimal combination of settings outperforming all other combinations. As of late, this has been a decisive turning point for computational stylistics, since it optimizes mere selection to 'selection by exclusion': if one roughly tries out a wide array of settings, one can use evaluation metrics (A.3.5, p. 342) to be more confident as to which set of parameters performs best. It also avoids the problem of cherry-picking: selecting whichever answer confirms the analyst's intuition.

[29] All technical details of this process have been left out of this chapter. For more information, see the appendix on pp. 322ff.

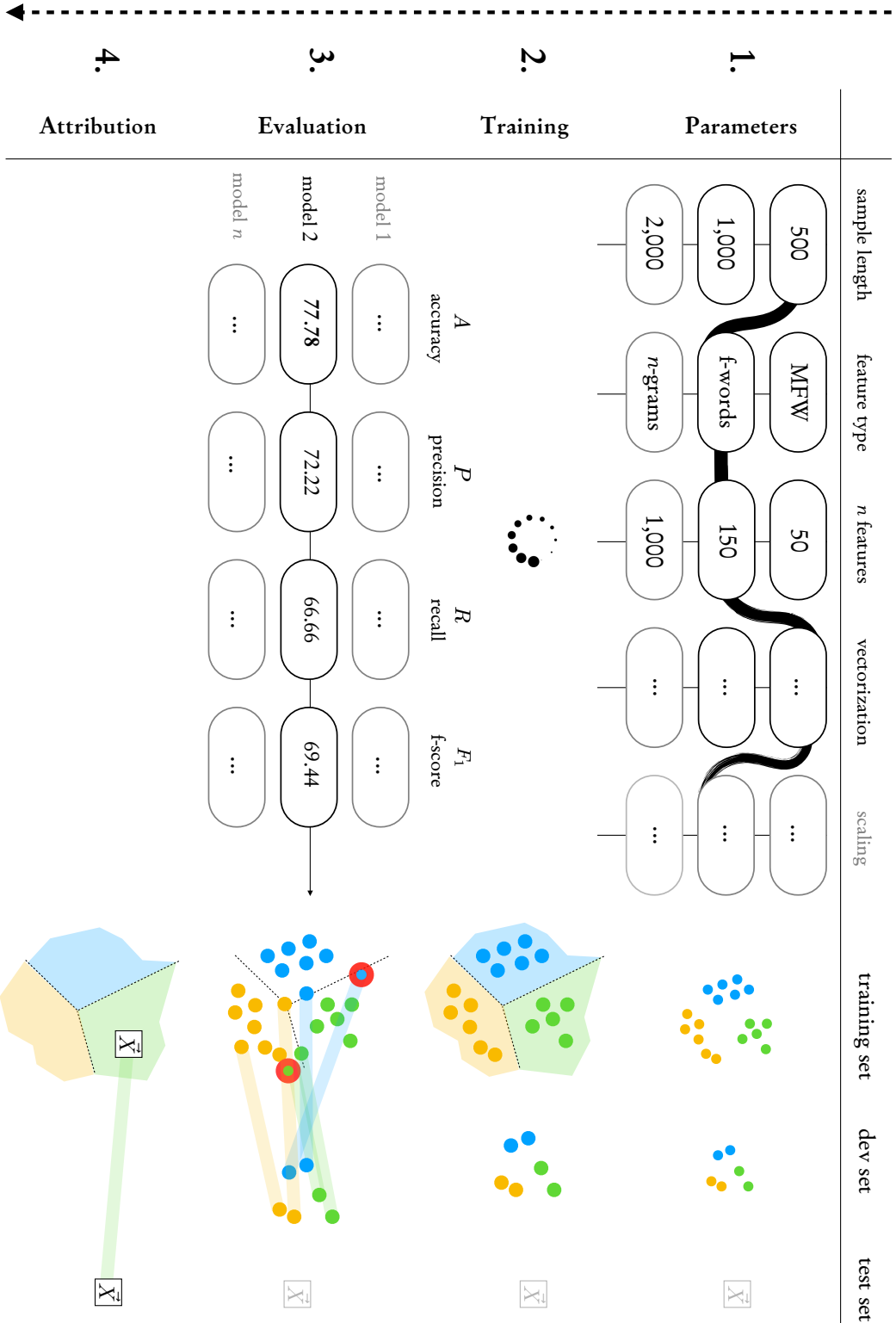



Figure 3.10: Step-by-step intuition of the training process of a hypothetical classifier.

### A Step-by-step Intuition

Fig. 3.10 provides a step-by-step intuition of how a machine-learning training process can be of assistance when trying to find the ideal settings for segregating works of different authorship. The gist of such a process (visualized in four steps on p. 108 from top to bottom) is the following:

**Step 1** Different parameters (or settings) are lined up in order to be tried in all combinations. The examples of the parameters in fig. 3.10 repeat those on the previous pages and figures, although the number of parameters one decides to search can be practically endless. The visualizations on the right end show how the data is partitioned into a training set, a development set and a test set. The test set in the far right column is an actual anonymous (test) text, and is left out of the training process at all times. The texts of known authorship, however, are divided into a training and development partition (often encountered as 'dev' set), where the training set outsizes the development set (e.g. by a proportion of 80% against 20%). During the model's actual learning process, only the training set is included, whereas the dev set is temporarily withdrawn from the experiment. After training, the dev set is allowed back in the game, and serves to evaluate if the model's predictions are sufficiently accurate or not. This is a way of 'assessing' if and how the model should 'develop,' hence, development set. In reality, this whole process is a tad more complex, and is also accompanied by so-called cross-validation methods, but this can be safely ignored for now.[30]

**Step 2** Consequently, training takes place on the training set. There are as many models as there are parameters. The visualization on the right end demonstrates how a classifier is learning to demarcate the clusters of the pre-labelled samples, and is drawing a decision boundary that sets them apart. The background colours suggest how the classifier associates certain regions ('vector spaces,' see p. 334) with the stylistic behaviour of particular authors.

**Step 3** After the model has learned to distinguish the texts in the training set, the dev set is introduced to give an indication of how well the model would perform when it was confronted with material 'in the wild.' Evaluation metrics such as accuracy, precision, recall and f1-scores (p. 342) help to assess all trained models, and if so desired the analyst can now build up a ranking of which models perform better and which perform worse.

**Step 4** Final attribution of the test text takes place after the dominant model has been chosen.

[30] For a full discussion on the division of train, development and test material, see p. 337 in the appendix. Cross-validation is explained on p. 340.

### Comparing Evaluations and Ranking Them

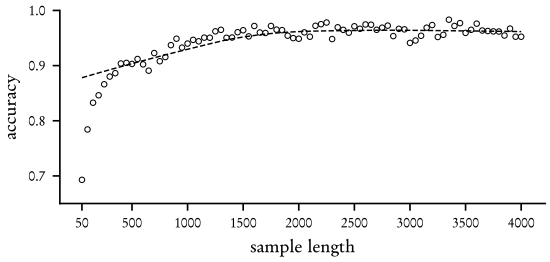
Step 4 provides some relief in the sense that it is possible to compare a range of different settings on an equal footing instead of having to randomly select whichever set of options works out. Not only does it allow to predict the behaviour of a certain set of parameters, it also allows to select one or more models that appear best equipped to tackle a particular classification problem at hand. Quite often, plotting how the behaviour of these models is subjected to the analyst's modifications can be instructive as to which are the crucial points at which they differ, and which parameters should be tended to or tweaked.[31] This also hauls much of the machine-learning process out of the black box and renders a model's gradual learning process intuitive.

To further illustrate this point, we return to one of the test cases in section 3.3.1 on the multiplicity of result and subject them to some of the machine-learning techniques just discussed. As one will recall, the four PCA plots had illustrated that different parameters invited varying results and therefore varying interpretations. Fig. 3.11 visualizes how this 'changeability' does not necessarily present an impasse, as each alternation can be gauged for its positive or negative effect. The figure contains an exhaustive search across different sample length settings for an increasingly larger set of most frequent words (MFW). Each plot represents a different number of features (75, 150, 400 and 1,000), and in each plot the classification accuracies for increasing sample lengths are presented. The experiment was run for the corpora of Guibert of Nogent, Peter of Celle and William of Saint-Thierry (in fig. 3.7). For each plot, accuracy of performance (y) is plotted in function of sample length (x). A series of linear SVM classifiers (p. 336) were trained for different sample lengths (x-axis), and consequently their performance was plotted (y-axis). When the lowess line levels, a saturation point for accuracy is found. This is the point at which a model is expected to reach an optimal trade-off between minimum sample length and maximum performance.

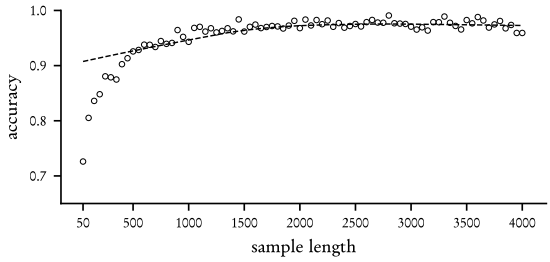
Plots such as these can teach the analyst all sorts of interesting information to take into account. An evident example is that for short samples, the accuracy of a model is generally low. As the samples increase in size, so does the classifier's performance, to the extent that the SVM can now guarantee by a percentage of approximately 95% (and an error margin of 5%) that it will correctly attribute an anonymous sample by one of these three authors 'in the wild.' That a steady point of saturation is generally reached around 2,500 words largely confirms former studies on sample length for Latin by Eder and Kestemont.[32] Furthermore, it is conspicuous that the four plots

[31] On a more technical side, it also provides illustrative means of deciding a trade-off between model complexity and computational cost.

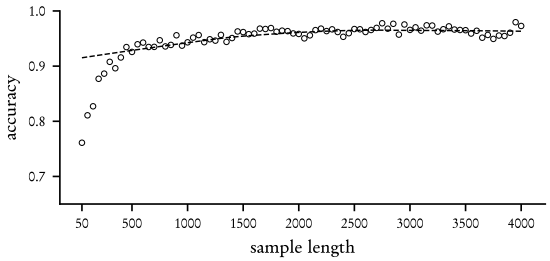
[32] Eder, "Does Size Matter?," see fig. 5 at 172 for Latin prose. Eder also stressed that his results were very much dependent on which corpus was used; and Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," see especially 211 and 214 (fig. 6), where Leave-One-Out cross-validation using Delta was performed. The corpus (Hildegard of Bingen's *epistolarium*) was smaller in size and the accuracy achieved was even higher as it rised above 95% when sample sizes were larger than 1,500 lemmas.



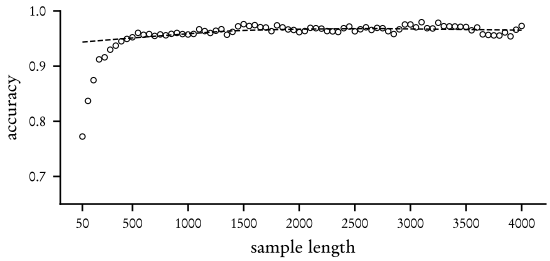

(a) 75 features. $\mu$ accuracy = 0.942. $\sigma^2$ (var.) = 0.0020.




(b) 150 features. $\mu$ accuracy = 0.955. $\sigma^2$ (var.) = 0.0019.




(c) 400 features. $\mu$ accuracy = 0.949. $\sigma^2$ (var.) = 0.0013.




(d) 1,000 features. $\mu$ accuracy = 0.958. $\sigma^2$ (var.) = 0.0009.

Figure 3.11: Cross-validated linear SVM classification accuracies (and fitted lowess line) for standard-scaled MFW feature vectors on the works of Guibert of Nogent, Peter of Celle and William of Saint-Thierry (as depicted in the PCA plots of fig. 3.7). The sample length was iteratively increased with 50 words per run. The data was partitioned into 5 (=$k$) folds. 5 times, the SVM was trained on 4 folds (training set), and tested on 1 held-out development set. The accuracy score is a result of the SVM's performance after each fold was used as a development set once. The folds were stratified (i.e. the data's original proportions were respected, see p. 341). The $c$ parameter was set to 1 (explained further below and on p. 336).

look rather similar. It appears that longer sample length has a much more visible positive effect on performance than longer vector length (not to be confused with each other).[33] Moreover, as is emphasized by the variance ($\sigma^2$) given in each of the figure descriptions, shorter feature vectors tend to output more inconsistent scores and behave more unpredictably, whereas longer feature vectors rise to their saturation point quicker.[34]

[33] The shortest feature vectors (75 features) are found in fig. 3.11a, as opposed to the longest (1,000 features) in fig. 3.11d. Adding additional features (from lower-frequency strata) than the first 75 MFW (higher-frequency stratum) yields no significant increase of performance, which can be deduced from the average ($\mu$) accuracy scores given underneath each of the 4 plots. On average, the accuracy for feature vectors of 1,000 features has risen only a percent and a half as opposed to 75-feature vectors (from 94.3% to 95.8 %).

[34] When using 75 features, the sample length saturates only from around 1,500 to 2,000 words, whereas with longer feature vectors, a sample of about 500 words already reaches a maximum accuracy of approximately 95%. Lengthier vectors appear to be more stable and boost the accuracy of very short samples. Even in cases where the classifier is expected to perform poorly in general —for example when the samples are only 50 words long— longer feature vectors provide more stable results. The accuracy percentage of 75-feature vectors for 50-word samples is 69.3%, whereas for 1,000-feature vectors the percentage has risen to 77.2% for that same sample size.

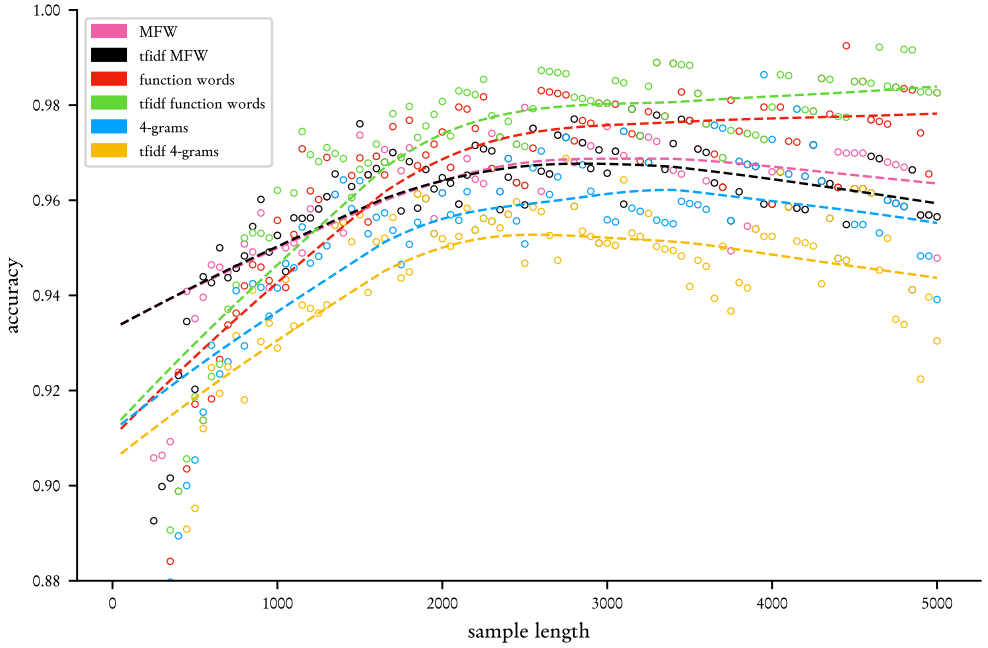



Figure 3.12: Cross-validated linear SVM classification accuracies (and fitted lowess line) for different feature vector types (see legend) on the works of Guibert of Nogent, Peter of Celle and William of Saint-Thierry (as depicted in the PCA plots of fig. 3.7). Number of features was set to 500. Standard scaling was always applied. The sample length was iteratively increased with 50 words per run. The data set was partitioned into 5 (=$k$) folds. 5 times, the SVM was trained on 4 folds (training set), and tested on 1 held-out development set. The accuracy score is a result of the SVM's performance after each fold was used as a development set once. The $c$ parameter was set to 1.

Many more settings than sample and vector length can be considered, and tested for their performance. There are as many classifiers as there are settings. Mapping out the inequality between these classifiers is what brings order to the chaos. A hierarchy needs to be established as to which number of features, sample lengths, feature types, ... etc. work best in which situations. For instance, the plots in 3.11 above only render the accuracy for one type of feature: MFW. One could decide to plot the various SVM performances of additional feature vector types (e.g. 4-grams, most frequent words or function words only), with the advantage of ascertaining which type of feature attains the best result at a chosen sample length. After all, the length of the anonymous test text is 'in the wild' seldom under the analyst's control. The text of unknown authorship could only be 500 words long, for instance. Moreover, it should not be surprising that for different data sets (which here means: different authors, different texts, different contexts), a different feature extraction method is better suited. Plotting the different performances of these feature types (still for the same corpus of fig. 3.7) in function of the sample length was done in fig. 3.12. The figure illustrates that if the test text should be under 1,200 words long, chances of correct attribution are higher when the feature extraction method is set to tfidf-weighted most-frequent words. After this threshold, tfidf-weighted function words take over the lead and achieve an increasingly

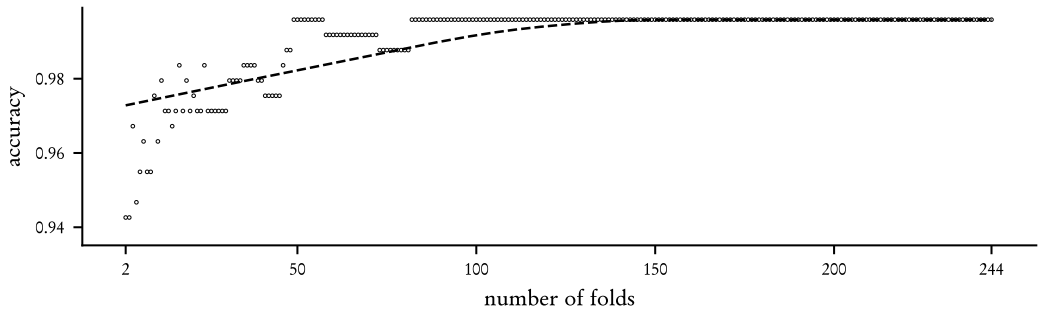



Figure 3.13: Cross-validated linear SVM classification accuracies for increasing number of folds (*k*) on the works of Guibert of Nogent, Peter of Celle and William of Saint-Thierry (as depicted in the PCA plots of fig. 3.7). Number of features was set to 500. Standard scaling was always applied. The sample length was 2,500 words. The *c* parameter was set to 1.

high accuracy floating between 97%–98%.

In addition, with a classifier such as SVM, there are several learning and classifier-specific parameters one can take into account. Fig. 3.13 is another example functioning within the same spirit of finding a fine-tuned set of parameters. The plot illustrates at what point the number of folds chosen in the cross-validation process reaches a saturated accuracy. As is discussed in more detail in the appendix on p. 340, during the classifier's learning process the data is split up into *k* blocks. In turns, one of these *k* blocks will serve as a development set, whereas the other *k*-1 blocks will serve as the training set. What one sees illustrated is the gradual transition on the x-axis of a non-stratified *k*-fold cross validation method to a Leave-One-Out strategy. The technical details can safely be disregarded for now.[35] At the core of the matter is that a more technical parameter such as this one also weighs in on the attribution result.

Another such a parameter —for SVM specifically— is the *c* parameter, for instance, which was set to 1 in all of our former experiments (figs. 3.11, 3.12 and 3.13). The consequence of such a low *c* is that the classifier will be very relaxed (a soft margin), or in other terms will allow for a higher recall and lower precision. Also the kernel method is a typical SVM parameter, which has in our former experiments consistently been linear. Generally, the differences in performance between various kernels will rarely be very high. Yet, even though linear kernel methods have been proven to work particularly well for most attribution problems,[36] it might sometimes be worthwhile

[35] The minimum on the x-axis of fig. 3.13 of course always equals at least *k*=2 folds. *K*=2 corresponds to splitting the training data in half and testing one half on the other and vice versa. Choosing the maximum number of folds, however, implies that the data is partitioned in as many folds as there are training instances, which corresponds to Leave-One-Out. One sees in fig. 3.13 that from around *k*=85 the algorithm has reached a saturation point, and becomes quite stable. Fewer folds (< 85) render the results more capricious. SVM models that were trained on fewer folds either score lower in accuracy, or else exhibit coincidental and unrealistically high assessments of their performance. From 85 folds onwards, no more significant impact on the algorithm's decision-making can be spotted, indicating that we have possibly reached the limit of the classifier's ability with the current settings. The attribution score, then, strands at an average of roughly 99.6% (between *k*=85 and *k*=244).

[36] Joachim Diederich et al., "Authorship Attribution with Support Vector Machines," *Applied Intelligence* 19 (2003):

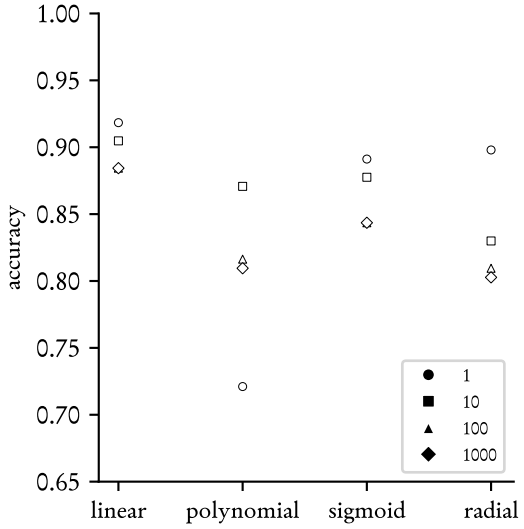



Figure 3.14: Cross-validated SVM accuracies for different *c* values and kernel methods. The folds were set to *k*=110. The markers indicate the *c* value applied.

pursuing if the linear method is not outperformed by non-linear extensions (or shapes), such as the polynomial, sigmoid or radial basis function.[37] Now that we know that feature vectors with 500 most-frequent words work quite well when the sample length is 4,000 words in this particular corpus (fig. 3.12) and that the model should be able to achieve a stable performance when the data is split into some 110 folds (fig. 3.13), we can delve into these classifier-specific hyperparameters such as parameter *c* and the various kernel methods.

Fig. 3.14 illustrates the accuracy score (y-axis) and its dependency on the kernel method used (x-axis). The symbol (clarified in the figure's legend) represents which value for *c* was used, varying from a soft (*c*=1) to a hard margin (*c*=1,000). As was to be expected, the linear kernel method with a soft margin (*c*=1) outperforms other kernel methods by a fairly big head start. This confirms the state of the art stating that linear SVM classifiers are superior for attribution problems.

## 3.4 Authorship Verification

### 3.4.1 The Confidence of Attribution

We have arrived at a point in this chapter where all sorts of statistical answers have been formulated to 'minimize doubt' over a statistical attribution of a medieval text. One expression of doubt has, however, for the present been left unmentioned. And

109–23.

[37] Ian H. Witten et al., *Data Mining. Practical Machine Learning Tools and Techniques*, 4th ed., The Morgan Kaufmann Series in Data Management Systems (Cambridge, MA: Morgan Kaufmann Publishers, 2017 (2000)), 255.

yet, it hinges on the absolute Ur-question of authorship attribution: what if our author is simply not amongst the included candidates? What if the author was an unlikely or unknown author, of whom no comparison texts have been transmitted? In most attribution problems, including those which we will be interested in in the current thesis, this is a realistic scenario.

A classifier's output is only reliable under the condition that the candidate is included in the data set. Imagine having a pupil learn the difference between apples, pears and bananas, before presenting it a tomato. Although the fruit appears unfamiliar, its features —round shape and red colour, to name but two— will lead a pupil to the conclusion that the tomato is an apple. Little can (s)he be blamed, since that is his or her limited frame of reference. Taking this limited experience into consideration, we have good cause to doubt the eventual prediction. In order to avoid these situations, a precautionary, built-in step in the experimental set-up needs to be devised, in which the classifier's adequate response would be to refrain from attribution and signalize the difficulties it is encountering. In other words, the classification procedure needs to be taught to doubt its own usefulness, and output an accompanying confidence score for the attribution it has made.

These intuitions fit in the philosophy of authorship verification, which may arguably be thought of as the ultimate extrapolation of the 'consensus'-idea advocated for in the previous paragraphs. Authorship verification methods tend to search many different parameters and features, and perform a same classification experiment over and over in order to convince themselves that the stylistic affinity between two documents is not based on 'superficial' but on 'deep-seated' similarities. They are arguably the smartest systems for authorship detection in the state of the art, and one of them, the *impostors* method (to be explained below), will have a prominent place throughout this thesis.

### 3.4.2 From One-Class Problem to Many-Candidates Problem

Authorship verification is what one may call a one-class problem.[38] We have a candidate author in mind (i.e. our 'class'), and we intend to verify whether or not an anonymous or pseudonymous document belongs to this class or not. The verification problem is to say the least challenging, and currently unresolved. With good reason. Koppel et al. pointed out that the verification problem is not just a variant or subtask of computational authorship attribution. It is the fundamental problem of the field.[39] Solving it equals finding the holy grail of all computational authorship detection, since every single attribution problem "can be decomposed into a series of

[38] Moshe Koppel and Jonathan Schler, "Authorship Verification as a One-Class Classification Problem," in *Proceedings of the 21st International Conference on Machine Learning* (New York: ACM Press, July 2004), s.p.

[39] Moshe Koppel et al., "The "Fundamental Problem" of Authorship Attribution," *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 284–91.

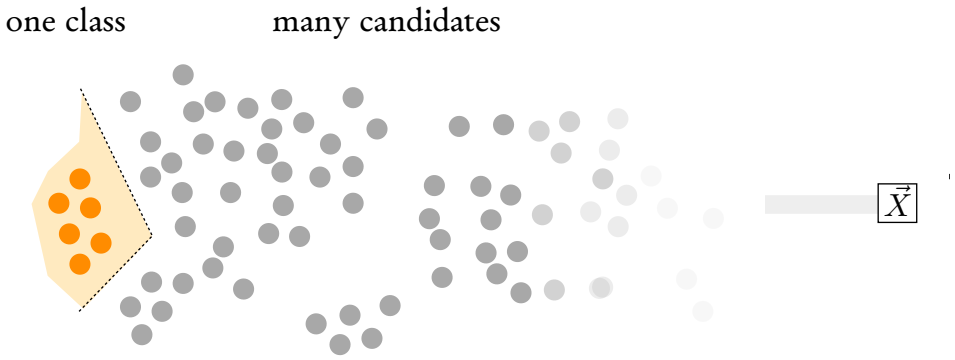



Figure 3.15: Intuition of the one-class problem reformulated as many-candidates problem.

verification problems."[40]

Verification can be rephrased as follows: we want to determine if two documents belong to one and the same class (i.e. are of same authorship).[41] Koppel and Winter were quick to point out that this 'one-class problem' virtually equals a 'many-candidates problem.'[42] An intuition of why this is the case is illustrated in fig. 3.15. Let us say, for instance, that we have an anonymous document, and we want to know if that document is written either 1. by Bernard of Clairvaux, or 2. not by Bernard of Clairvaux, meaning: any other author. In this example, Bernard of Clairvaux corresponds to our 'one class,' and any other author to 'many candidates.'

Ideally, 'many candidates' literally corresponds to an undetermined and infinite corpus of authors that have ever written Latin text. Having such a corpus increases the model's frame of reference. The inclusion of a wide background corpus renders the set-up its 'open' character as opposed to its 'closed' variants such as in the SVM experiments described in the previous section.[43] However, having an infinitely stretched corpus of authors at disposal is an unrealistic demand in practice, which explains why the authorship verification problem has as yet not been adequately handled. The background corpus, which ideally should contain 'all that has ever been written,' can only be approximated. This is where the benchmark corpus, by which we opened this chapter, gains a very practical purpose.

[40] Nektaria Potha and Efstathios Stamatatos, "An Improved *Impostors* Method for Authorship Verification," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, ed. Gareth J.F. Jones et al. (Dublin: Springer International Publishing, 2017), 138.

[41] This is essentially the title of the article in which the *impostors* method was first featured, see Moshe Koppel and Yaron Winter, "Determining If Two Documents Are Written by the Same Author," *Journal of the Association for Information Science and Technology* 65, no. 1 (2014): 178–87.

[42] Ibid., 180.

[43] Ibid.

### 3.4.3 The *Impostors* Method

#### Experimental Design

The *impostors* method is an intuitive, unsupervised attribution technique.[44] It uses similarity-based methods (A.3.5, p. 333), and it addresses very short texts (500 words) specifically. Its novelty is two-fold. As announced in the foregoing paragraph, the *impostors* method defends itself from coincidental attributions by relying on a background dataset from which it can select 'impostors.'[45] The impostors can be selected in a variety of ways, but most often they are chosen on the basis of strong lexical overlap, so that the problem is considerably hardened. One can think of the impostors as 'lookalikes,' documents by other authors that show a strong lexical resemblance.[46] This collection of impostors intend to disorientate the classification experiment, and constitute a surrogate —and very obnoxious— version of what we described above as the infinite class containing 'all that has ever been written.' In this thesis, the impostors will be drawn from the twelfth-century benchmark corpus of Latin authors (pp. 304ff.) The second novelty is that the *impostors* method consequently selects feature subsets (50%) from an original feature set (for Winter and Koppel it consisted of 100,000 words): "[...] if a particular candidate [...]'s known text is more similar to the snippet than any other candidate for many different feature set representations of the texts, then that candidate is very likely the author of the snippet."[47] The original feature background set consisted of some 100,000 features, from which randomly feature subsets (about half the size of the original background set) were selected in $k$ iterations.

The idea, similar to bootstrapping,[48] is that two documents might well attain a degree of similarity in some representation or another, but such a direct, first-order

[44] The *impostors* method was first introduced five years ago (in 2014) by Koppel and Winter. See Koppel and Winter, "Determining." In many aspects, the *impostors* method constituted an answer to some of the criticisms voiced towards the *unmasking* method, another popular authorship verification technique, introduced to the field some ten years earlier in 2004. A portion of the appendix in this thesis is dedicated to explaining *unmasking*, see pp. 351ff. For the article on author *unmasking*, see Koppel and Schler, "Authorship Verification as a One-Class Classification Problem." Although the simple ingenuity of *unmasking* had been well appreciated in the field, there were some minor setbacks to the method, such as the heavy reliance on supervised and computationally costly methods (the intensive re-training of a classifier), its cumbersome implementation, and reported failures in cases of shorter sample lengths (below 10,000 words), see for instance Conrad Sanderson and Simon Guenter, "Short Text Authorship Attribution via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)* (Sydney, July 2006), 482–491.

[45] Thereby proposing an "approach [...] based on the solution to a closely related problem: Given a large set of candidate authors, determine which, if any, of them is the author of a given anonymous document," in Koppel and Winter, "Determining," 180.

[46] The lexical overlap should of course be limited to mid-frequency items, and should best not take the top-frequency items such as function words into account. Querying impostors by using a small subset of words is called the 'On-the-fly' method for selecting impostors, see ibid., 182.

[47] As features, the method's founders Koppel and Winter proposed space-free character 4-grams vectorized to Tfidf values (Term frequency inverse document frequency. See p. 321). See ibid., 181.

[48] Stover and Kestemont, "The Authorship of the *Historia Augusta*," 146.

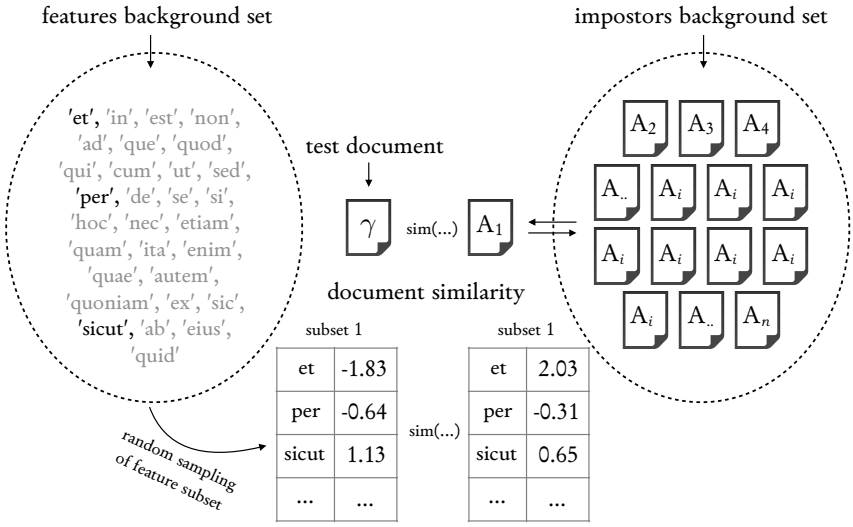



Figure 3.16: The *impostors* method visualized. The above process is repeated $k$ times. On the left one finds the original features background set from which subsets are stochastically sampled, on the right one finds the impostors background set.

similarity is insufficient to establish certainty. In order to prove consistent similarity and attribute the text, the match between two documents needs to be validated across $k$ runs (Winter and Koppel proposed 100 runs) with varying feature subsets. Ultimately, for some test text ($\gamma$ in the figure), a ranking for all candidates in the background set $< A_i ... A_n >$ is proposed. The proportion of times that some candidate is the top match is consequently tested against a benchmark threshold $\sigma^*$, which, if not surpassed, lets the impostors algorithm output the category label "None of the above." For instance, if 84 out of 100 iterations some candidate author $A_8$ was the top match for the test text, the score would be $\frac{84}{100}$. Given that the threshold was set at 0.80 ($\sigma^*$), $A_8$ can be considered author of $\gamma$ by a very high probability. Anything below that threshold outputs 'Uncertain.'

The threshold $\sigma^*$, denoting the proportion of required document matches before valid attribution, can of course not be chosen at random. Finding such a threshold is ideally based on former experience, hence the value of a pre-labelled training corpus. The *impostors* method relies on learned instances of known labelled pairs of <*same-author*> or <*different-author*> as training data.[49] For both types of classes we

[49] One can compare this threshold to a classification threshold, explained in the appendix on p. 345. The main idea in the training process is to benchmark <*same-author*> and <*different-author*> sets (in equal number) against impostor documents that are randomly sampled from the benchmark corpus. One will find in the literature that impostors are sometimes chosen by their close lexical overlap to test text $\gamma$, which means that the impostors are very convincing. This aspect of 'impostor quality' was suggested in Koppel and Winter, "Determining." To our own experience, this works less well than random sampling. We therefore chose the method as applied in Kestemont et al., "Authenticating the Writings of Julius Caesar," see especially 88. In other words, we feed the model a text sample $\gamma$ in turn, as in fig. 3.16, and benchmark it against a number of documents in which we either involve the true author (<*same-author*>) or not (<*diff-author*>). For both class labels we collect percentages indicating how many times out of $k$=100 times correct and false attributions were made.

ultimately collect percentages indicating how many times out of $k$=100 times correct and false attributions were made. This mean percentage can consequently function as our learned threshold.[50]

### In Practice: The *Impostors* Method and the Benchmark Corpus

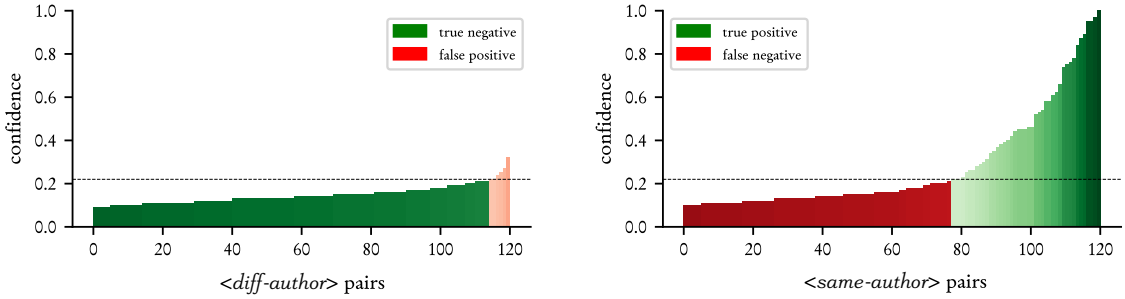
Fig. 3.17 gives an intuition of the thresholding mechanism at work in the *impostors* method. These specimen specifically were in fact taken from the training process that preceded the rolling *impostors* method for the *Vita Hildegardis* in chapter 6 (see pp. 188ff.). The *impostors* method can be a very powerful and meticulous method when it trains well on the authors under scrutiny. If a well-balanced $\sigma^*$ can be established, this grants great confidence to any test attribution following it, especially if this attribution has a firm and high confidence score. On the other hand, as noted earlier, authorship verification remains an extremely difficult and unsolved problem, exploring the limits of what is feasible in the current-day landscape of computational stylistics. That is, for instance, why it has up until today remained one of the main challenges of the annual PAN competition.[51]

Much more methodological research needs to be done over whether or not this type of method works well for historical texts written in medieval Latin, especially for authors who borrow and exchange texts as extensively as in the current corpus. Although there are a number of parameters to explore, such as the number of impostors, the degree of how convincing these impostors are as opposed to test text $\gamma$, the feature subset size, the feature type, etc. the great weakness of the *impostors* model is that much depends on the training of a single parameter, threshold $\sigma^*$. A common problem is that to attain high accuracy and precision, a considerably low recall needs to be allowed for (as becomes clear by the number of false negatives in subfig. 3.17b). Sometimes, finding such a balance even proves to be impossible altogether. In some situations the *impostors* method is extremely precautious in making attributions and ultimately quite worthless, as it refrains from taking any risk and attributing the text to any of the candidates at all (many false negatives). On other occasions the model is more relaxed, but also makes more errors (many false positives). My impression is that this mostly has its cause in data itself, which is indeed typified by a considerable amount of lexical overlap and exchange between authors. I will therefore sometimes choose to borrow

Ideally, $\sigma^*$ will be low for <*diff-author*> pairs, and high for <*same-author*> pairs. Adjusting this threshold establishes a trade-off between precision (of all selected samples, how many were correct) and recall (of all correct samples, how many were selected). These evaluation metrics are discussed more elaborately on pp. 342ff.

[50] Such a threshold can be surprisingly stable. Stover et al., for instance, in applying the *impostors* method for the *Compendiosa expositio* by Apuleius, reported of 199 <*different-author*> pairs, where only three pairs obtained a score above a $\sigma^*$ threshold of .20. Of the 32 <*same-author*> pairs, 15 scored above the $\sigma^*$ of .50. Consequently, one can use these observations to find an ideal trade-off between high precision (fewer acceptances, fewer errors) and high recall (more acceptances, more errors) for $\sigma^*$ (A.3.5, p. 342). See Stover et al., "Computational Authorship Verification Method," especially 241.

[51] See p. 11 in this thesis, and Stamatatos et al., "Overview of PAN 2018."




(a) Training on <*diff-author*> pairs.

(b) Training on <*same-author*> pairs.

Figure 3.17: Intuition of *impostors* method thresholding. The y-axis indicates the confidence by which the attribution was made, whereas the x-axis indicates the sample index number (there were 120 samples in total). Note that the number of samples per class (<*same-author*> and <*diff-author*>) is equal. The trained threshold $\sigma^*$ is visualized by the horizontal, dotted line. Its purpose is to avoid as many false decisions as possible (red) in favour of correct ones (green).

principles from the standard *impostors* method, but propose variations that relax the strict $\sigma^*$ condition, as a second-best alternative to a method that would theoretically be superior, but is in practice not always feasible. Poor performance during training forced me to step down the 'open' experimental set-up to a 'closed' many-candidates attribution problem, one that might not be a perfect approach of a situation 'in the wild,' but highly successful in making closed-set attributions as opposed to a great number of 'distractors' (instead of 'impostors').

## 3.5 Conclusive Remarks

This chapter has been quite a long journey, guiding us past a wide range of methods. In the previous pages, the elementary technical aspects of computational stylistics have been explored, as have they been put into practice so as to gain an acquaintance with the method's strengths and weaknesses. Now that we have come to the close of this chapter (and part 1 of this thesis), a concise summarizing statement to frame chapter 3's significance in the larger picture is in order.

One lead notion, which has been running through all the different subsections above, deserves due attention: the establishment of consensus. It has become clear that the style of an author can (and should) be measured in a variety of ways. Caution is advised in situations where only a singular result is generated and presented as final. Be it in the selection of style features, in the selection of parameters, in the selection of models, etc., it is sound practice to search them exhaustively, evaluate and compare the outcomes, and —if possible— ultimately consent upon a method that works best for a given problem. The current state of the art, then, is well served by bootstrapping-inspired methods, in which hundreds and thousands of tasks are performed which automatically generate a consensus. That is why the *impostors* method was presented

as last above, and will play an important role throughout most of the case studies. In many ways it forms the ultimate extrapolation of the idea of consensus, and was therefore termed the 'fundamental' problem of authorship attribution.[52] Its ability to harden the problem by involving a world 'larger' than a given set of candidates and thereby imitating the practical reality of how a text of doubtful authorship is encountered 'in the wild,' is on a theoretical level at least a promising avenue for research into the computational detection of authorship. It might bring us yet a step closer to the distributional author.

[52] Moshe Koppel et al., "The "Fundamental Problem" of Authorship Attribution," *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 284–91.

# Part 2

# 4

# Scandal in Clairvaux: Bernard and Nicholas of Clairvaux*

## 4.1 Bernard of Clairvaux and his Secretaries

Despite the fact that he is so well-known, Bernard of Clairvaux (c. 1090-1153) is an elusive figure. As the prominent Bernard specialist Dom Jean Leclercq pointed out: "there always have been, as it were, several Bernards."[1] Especially after Bernard's lifetime, myth and legend contributed richly to his prismatic conception, not always complementary or at times even contradictory to what historical evidence suggests. Such a multiplicity of Bernards has often been found troublesome. The Cistercian abbot is notoriously hard to pin down. This issue manifests itself just as acutely in his literary heritage, which at times also betrays the presence of "several Bernards." As an icon and figurehead of the Cistercian movement, Bernard was constant subject to imitation, which is only partly explained by his spiritual authority. Of equal importance was his skill in literary composition. Bernard's style was of such grandeur that it was imitated by the greatest theologians of his time. He provided an architecture for a Cistercian way of writing. For example, Aelred, the abbot of Cistercian daughter house Rievaulx (c. 1110–67), came to be called "the English Bernard."[2] But Bernard's very best imitators were in fact found by his side, in the scriptorium of Clairvaux.

By 1145, the abbot's acclaim as the icon and figurehead of the Cistercian movement

* The case study treated in this chapter has been formerly published as De Gussem, "Bernard of Clairvaux and Nicholas of Montiéramey." This publication was reintegrated in this thesis with some minor revisions and additional experiments. The data treated both there and in this chapter was rearranged according to chronological principles. To inspect, download and reuse the data, go to the aforementioned article's repository at `https://github.com/jedgusse/bernard`.

[1] Jean Leclercq, "Toward a Sociological Interpretation of the Various Saint Bernards," in *Bernardus Magister: Papers Presented at the Nonacentenary Celebation of the Birth of Saint Bernard of Clairvaux, Kalamazoo, Michigan; sponsored by the Institute of Cistercian Studies, Western Michigan Univ., 10-13 May 1990*, ed. John Robert Sommerfeldt, Cistercian Studies Series 135 (Spencer: Cistercian Publications, 1992), 19–33.

[2] William Michael Ducey, "St. Ailred of Rievaulx and the *Speculum Caritatis*," *The Catholic Historical Review* 17, no. 3 (1931): 308.

had brought along such a considerable administrative workload that the assistance of a group of secretaries —which, it could be argued, amounted to a kind of chancery— was indispensable.[3] These secretaries acted as Bernard's stand-ins and spared him the time and effort it would have cost if he had to take up the quill himself at every single occasion.[4] The *reportatio* entailed that the contents of Bernard's letters or sermons be engraved on wax tablets in a tachygraphic fashion. The cues, keywords, and biblical references that Bernard had spoken aloud provided a framework that captured the gist of his diction.[5] Afterwards, the scribe reconstructed what he had heard as a text on parchment, which could pass for Bernard of Clairvaux's in its literary allure. Among these amanuenses, Nicholas of Montiéramey († 1176/78) was a focal figure and a highly skilled imitator of his master's writing style. The influence of Nicholas's mediation on several particular texts within Bernard's corpus, and more generally on his entire oeuvre, has been subject to much debate. In this chapter we revisit the authorship of a selection of texts from Bernard's corpus. Generally, this corpus comprises Nicholas of Montiéramey's letters and sermons[6] and Bernard of Clairvaux's letter corpus (*Corpus epistolarum*), *Sermones de diversis* (hereafter *De diversis*), and *Sermones super Cantica Canticorum*.[7]

## 4.2 Nicholas of Montiéramey

### 4.2.1 Biographical Details

The daily routines and workings of Clairvaux's chancery are rather poorly documented. We rarely know any of the scribes by name, and for those whom we do —a select group of six— only three give us a faint clue of their specific tasks and responsibilities.[8] Nicholas began serving Bernard as an emissary around 1138–41, car-

[3] Some might argue that the word 'chancery' is inappropriately used of Bernard's scriptorium, as it was not primarily a formal or institutional body of administration charged with the composition and dispatch of official documents.

[4] The workings of Clairvaux's scriptorium are extensively investigated in Rassow, "Die Kanzlei Bernhards von Clairvaux"; and Jean Leclercq, "Saint Bernard et ses secrétaires," in *Recueil d'études sur saint Bernard et ses écrits*, vol. 1, Storia e Letteratura 92 (Rome: Edizioni di Storia e Letteratura, 1962), 3–25; Constable also commented on the difficulty of the redaction process: "Aside from a few outlines dictated by Bernard or based on sermons he gave, most of the surviving texts are later compositions drawn up by either himself or his secretaries, and they bear little resemblance to what he actually preached, if they were ever delivered," in Constable, "Language of Preaching," 134.

[5] Stenography, or shorthand systems, had been forgotten by the twelfth century, making place for tachygraphy, a rapid form of writing: see Parkes, "Tachygraphy in the Middle Ages."

[6] Nicholas of Montiéramey's letters can be found under *PL* 196:1593–1651. The sermons have been identified in Jean Leclercq, "Les collections de sermons de Nicolas de Clairvaux," in *Recueil d'études sur saint Bernard et ses écrits*, vol. 1, Storia e Letteratura 92 (Rome: Edizioni di Storia e Letteratura, 1962), they are collected among those of Peter Damian in *PL* 144. For a comprehensive list of these sermons (by title), see the appendix in A.4.1.

[7] These works are edited in Bernardus Claraevallensis, *Sancti Bernardi Opera*, ed. Jean Leclercq, Henri M. Rochais, and Charles Holwell Talbot (Rome: Editiones Cistercienses, 1957–77), vols. 7–8 comprise the *Corpus epistolarum*, vol. 6 the *Sermones de diversis* and vols. 1–2 the *Sermones super Cantica Canticorum*.

[8] We know that Bernard's earliest secretary was William of Rievaulx. He must have been active from 1120 until 1132, before travelling to northern England to establish the monastery of Rievaulx in the diocese of York, a daughter house for

rying letters concerning Abelard's heresy to Rome. At this time he was still chaplain of Hato, bishop of Troyes, and Peter the Venerable's friend and secretary, but he must already have been collaborating with Bernard from 1140 onwards.[9] He would officially become a monk at Clairvaux around the end of 1145. His literary qualities, likely to have been acquired through his education in the Benedictine abbey of Montiéramey,[10] enabled him to enter the scriptorium immediately and officially become Bernard's closest secretary. He appears to have been responsible for supervising the workings of the chancery,[11] and he may have been the monastery's librarian.[12]

### 4.2.2 Scandal in Clairvaux

But their friendship knew an abrupt and painful ending in the final years of Bernard's life, around 1151–52, when Nicholas must have severely breached his master's trust. In a letter to Pope Eugene III, we find Bernard disconcerted over the fact that letters had been sent out under his name and seal by "false brethren" without his permission.[13] Later, Bernard would identify Nicholas as the culprit among these brethren,[14]

Clairvaux, to become its first abbot. See Rassow, "Die Kanzlei Bernhards von Clairvaux," 5. William's intimate bond with Bernard must have established a solid base upon which Clairvaux and Rievaulx were able to cooperate, communicate, and exchange recruits: see Brian Patrick McGuire, "Introduction: A Companion to Bernard of Clairvaux," in *A Companion to Bernard of Clairvaux*, ed. Brian Patrick McGuire, Brill's Companions to the Christian Tradition 25 (Leiden: Brill, 2011), 8. Three other names that have come down to us are Balduin of Pisa, Gerard of Peronne, and Raynaud of Foigny, but none of these seems to have had much significance: see Evans, *Bernard of Clairvaux*, 20. A more important personality was Geoffrey of Auxerre (c. 1115–c. 1188), who was a former student of Peter Abelard and allegedly denounced the Parisian schools in favour of the monastery after having witnessed Bernard's genius and eloquence in preaching: "continuo tres ex illis compuncti sunt et conversi ab inanibus studiis ad verae sapientiae cultum, abrenuntiantes saeculo et Dei famulo adhaerentes," in Gaufridus Autissiodorensis, *Libri III–VI Sancti Bernardi abbatis Claraevallensis vita et res gestae*, in *PL* 185:301–51, ed. Jean Mabillon, 327. He entered Clairvaux in 1140 and became the abbot's secretary in 1145, a time when the administrative obligations in Clairvaux reached their peak and an official chancery had been established. He would become abbot of Clairvaux himself in 1163 but had to abdicate his leadership after two years, presumably as a consequence of an internal dispute over the papal schism between Alexander III and Victor IV. See the introduction to *Goffredo di Auxerre: Super Apocalypsim*, ed. Ferruccio Gastaldelli, Temi e Testi 17 (Rome: Storia e Letteratura, 1970), 14–5.

[9] There is scholarly debate over when exactly Nicholas initiated his collaboration with Bernard, but recent research tends to agree that it must have been earlier than his accession in 1145–46. See Constable, "Dictators and Diplomats in the Eleventh and Twelfth Centuries: Medieval Epistolography and the Birth of Modern Bureaucracy," 43–4: "Nicholas was at Clairvaux probably from the early 1140s to 1152 and assisted Bernard with his sermons as well as his letters, but he continued to visit Cluny and to serve Peter the Venerable, one of whose letters, we have seen, he presented to Bernard orally"; and Turcan-Verkerk, "*Ars Dictaminis*," 70: "Ami de Pierre le Venerable, [Nicolas] avait déjà servi les intérêts de Bernard en portant au pape, en 1140–1141, des lettres concernant Abelard—à la rédaction desquelles il avait peut-être déjà participé, comme le suggère le manuscrit Phillipps 1732 [...]. Trois billets de recommandation envoyés par Bernard à Innocent II entre 1138 et 1143 semblent le concerner [Epp. 434–36], et montrent que s'il servait Hatton, il le faisait en obéissant à Bernard."

[10] "Nicolas fit ses études à l'abbaye bénédictine de Montiéramey, près de Troyes en Champagne. On parle souvent de lui comme d'un *Magister*,"see John F. Benton, "Nicolas de Clairvaux," in *Dictionnaire de Spiritualité*, vol. 11 (Paris: Beauchesne, 1982), 255.

[11] Jean Leclercq, "Lettres de S. Bernard: histoire ou littérature?," in *Recueil d'études sur saint Bernard et ses écrits*, vol. 4, Storia e Letteratura 167 (Rome: Edizioni di Storia e Letteratura, 1987), 148.

[12] Giles Constable, ed., *The Letters of Peter the Venerable*, Harvard Historical Studies 78 (London: Harvard University Press, 1967), 2:321.

[13] "Periclitati sumus in falsis fratribus," see Bernardus Claraevallensis, Ep. 284, *SBO* 8:198–9.

[14] In a letter likewise addressed to Eugene III, see Bernardus Claraevallensis, Ep. 298, *SBO* 8:214.

although the exact reasons why the latter deserved this accusation are nowhere explicitly disclosed.[15] In any case one can assume from his correspondence and his own words[16] that Nicholas's talent as a writer and his "versatility" ingratiated him with the greatest men of his time.[17] Equally so, Nicholas appears to have had —perhaps through this flamboyance and self-confidence— a talent for making enemies as well.[18]

The scandal at Clairvaux and the breach of Bernard's trust has for a long time upheld the portrayal of Nicholas as a disreputable Judas by Bernard's side, an analogy for which Bernard himself was responsible.[19] Conversely and simultaneously, Bernard's status as a saint continued to grow during the intense process of canonization and idealization following his death.[20] These respective caricatural depictions, in which Nicholas was deplored as the mistrusted secretary and Bernard praised as the saint who had become victim of textual theft, show through on an academic level as well. Jean Leclercq was as relentless as Bernard in accusing Nicholas of deceit, shamelessness, and plagiarism.[21] Nicholas's most striking example of seeming textual theft presents itself in his letter to Henry the Liberal, count of Champagne, to whom he humbly offered his services as a secretary shortly after his expulsion from Clairvaux. Accompanying the letter we find nineteen sermons originally attributed to Peter Damian,[22] nine sermons attributed to Bernard of Clairvaux,[23] and seventy-four short commentaries to the Psalms that are ascribed to Hugh of Saint-Victor (c. 1096–1141).[24] In the let-

[15] Constable, *The Letters of Peter the Venerable*, 327.

[16] Nicolaus Claraevallensis, *Epistolae*, in *PL* 196:1593–654, ed. Jean Mabillon, 1652: "Ab ineunte aetate mea placui magnis et summis principibus hujus mundi,"

[17] The word is Jean Mabillon's in *PL* 183:25: "Vir fuit ingenii facilis, versatilis, facile in aliorum affectus influens."

[18] An example can be found in Nicholas's dispute with Peter of Celle. The two "were at odds over a substantive matter, a theological point about how to treat the attributes of God, and Abbot Peter took offense that Nicholas, who should have possessed the power Nicholas was accused of 'inverting words and their meaning,' a characteristic which Peter interpreted as equal to a falsification of language: 'verba quoque et sensus verborum praesumis quandoque invertere,'" see Petrus Cellensis, *Epistolae*, in *PL* 202:405–635, ed. Ambroise Janvier, 512.

[19] Bernard literally made the analogy with Judas, which he significantly did not make often in his letters: see Brian Patrick McGuire, "Loyalty and Betrayal in Bernard of Clairvaux," in *Loyalty in the Middle Ages: Ideal and Practice of a Cross-Social Value*, ed. Jörg Sonntag and Coralie Zermatten (Turnhout: Brepols, 2015), 317–8.

[20] As it was first initiated by his biographer in Guillelmus de Sancto Theodorico, *Liber I Sancti Bernardi abbatis Claraevallensis vita et res gestae*, in *PL* 185:225–68, ed. Jean Mabillon, William of Saint-Thierry (1075–1148) wrote the first book of Bernard's *Vita*, a biography with a hagiographical, panegyrical slant. He shares the authorship of the entire *Vita* with Bernard's secretary Geoffrey of Auxerre (c. 1115–c. 1188) and the Benedictine abbot Arnaud de Bonneval. Geoffrey was a strong advocate for Bernard's canonization, in which William's texts played a fundamental role.

[21] Leclercq cannot but express his dislike for Nicholas in phrases such as "cet homme sans caractère, mais lettré, doué de mémoire, habile à manier les fiches, prompt à entrer 'dans le personnage' d'un autre, aurait pû être pour S. Bernard un parfait secrétaire, si seulement il avait été honnête," or "or la suite du recueil prouve qu'il était sans scrupules en ce domaine comme en d'autres," or "ainsi les témoignages les plus formels de Nicolas lui-même sont trompeurs, car il ment," see Leclercq, "Sermons de Nicolas de Clairvaux," 56–8; and elsewhere, in Jean Leclercq, "Deux épîtres de Saint Bernard et de son secrétaire," in *Recueil d'études sur saint Bernard et ses écrits*, vol. 2, Storia e Letteratura 104 (Rome: Edizioni di Storia e Letteratura, 1966), 317: "On sait combien cet esprit peu original aime se citer lui-même, reprendre, en les modifiant à peine, des expressions qu'il a déjà employées en d'autres écrits."

[22] See A.4.1, p. 354

[23] Now collected as sermons 6, 7, 21, 62, 83, 100 and 104 in Bernardus Claraevallensis, *SBO* 6.1.

[24] On the sermons, see Leclercq, "Sermons de Nicolas de Clairvaux," 57. Hugh's commentaries or chapters, the *Adnotationes elucidatoriae in quosdam Psalmos David* —which forms the second part of the *Miscellanea*— are collected under

ter, Nicholas asserts that these writings are "of my invention, of my style, aside from what I have taken from others in a few places."[25] We know this assertion to be true of the nineteen sermons also found among those of Peter Damian, which have been identified by Leclercq as stemming from Nicholas. Bernard's and Hugh's writings, on the other hand, appear to have been copied almost literally, not merely rearranged or paraphrased "in a few places" (*paucis in locis*), as Nicholas seems to suggest. Most of the nine sermons can be found in Bernard's *De diversis*. It is striking that, months after his banishment from Clairvaux, Nicholas seemingly betrays his former abbot again with what appears to be a willful appropriation of Bernard's texts.

Such incriminating evidence contributed to his reputation as a plagiarist, this reputation in its turn provoking prejudicial conclusions in other attribution issues. Henri M. Rochais, for instance, in a codicological approach to the question of determining the disputed authorship of three other lengthy sermons in Bernard's *De diversis* corpus (*De diversis* 40, 41, and 42),[26] pointed out these sermons' close similarities to two of Nicholas's works and to chapter 100 of Hugh of St. Victor's Victor's sixth book of *Miscellanea* (the well-known writer somehow seems to be involved again);[27] yet stated with confidence that Nicholas stole the texts from Bernard under false pretences. This hypothesis Rochais sees corroborated in "the secretary's unscrupulous personality."[28] At the same time Rochais casts aside Mabillon's belief that the literary style of these sermons hardly seems that of Bernard as an all-too-subjective and unscientific argument.[29] To our view, Rochais' own subjective mistake was that —despite being fully aware of Bernard's collaboration with his secretaries— he treated codicological unity as identical to stylistic or authorial unity: "Cette tradition manuscrite ne donne donc aucun motif de doute sur l'authenticité bernardine des trois sermons étudiés, et, au contraire, elle constitue une telle probabilité en faveur de cette authenticité, qu'il faudrait des arguments incontestables pour dénier à Bernard leur composition."[30]

Leclercq's and Rochais' attributions still stand in their editions, widely used to-

*PL* 177:589–634.

[25] My translation. The original Latin says "Meo sensu inventos, meo stylo dictatos, nisi quod paucis in locis de sensibus alienis accepi." The text is found in the prefatory letter in MS Harley 3073 and has been edited in Leclercq, *Recueil d'études* 1:49–50.

[26] Sermo 40, *De viis vitae quae sunt confessio et oboedientia*; sermo 41, *De via oboedientiae*, and sermo 42, *De quinque negotiationibus, et quique regionibus*. See Bernardus Claraevallensis, *SBO* 6.1:234-61.

[27] The specific text referred to is Hugh of Saint-Victor, *De septem gradibus confessionis*, in *PL* 177:856–58. Henri M. Rochais, "Saint Bernard est-il l'auteur des sermons 40, 41 et 42 «de diversis»?," *Revue Bénédictine* 72, nos. 1–2 (1962): 326. There is a lack of clarity as to how exactly Hugh of Saint-Victor's *Miscellanea* was constituted—whether the collection was assembled by Hugh himself or whether it is a compilation assembled from his writings by others.

[28] "Le caractère de ce secrétaire peu scrupuleux rend assez vraisemblable l'hypothèse d'un nouveau plagiat de Nicolas aux dépens de son ancien abbé," see ibid., 326.

[29] "Dom J. Leclercq a dit justement ce qu'il faut penser de cette sorte d'argument trop subjectif pour avoir, à lui seul, une valeur réellement probante," see ibid., 325. Mabillon's argument for attributing the sermons to Nicholas can be found in a note to *PL* 183:647–48: "Hic sermo sequensque in editione Lugdunensi anni 1514, in qua primum prodiere, extra classem genuinorum Bernardi sermonum locati sunt; nec stylum ejus plene assequi videntur."

[30] Ibid., 330.

day, although medievalists have seriously contested their highly subjective and speculative approach towards authorship attribution and their prejudiced view of Nicholas of Montiéramey's alleged deceitfulness and falsification. Nicholas's appropriation of some of his former master's works in his letter to Henry the Liberal is rather the continuation of a dialogue,[31] not a spiteful act of revenge. Stephen Jaeger has similarly argued that Nicholas indulges in the kind of imitatio that would have made little distinction between 'honest' and dishonest intentions.[32] Like any distinguished writer of his time, Nicholas carefully applied for a new position by showcasing his complete immersion in a prevalent literary network. After all, Nicholas was applying for a position as Henry's new secretary, and a familiarity with the greats of the twelfth century would have been one of the prerequisites. Leclercq's assertion that Henry the Liberal must not have noticed Nicholas's blatant plagiarism because he was a layperson unfamiliar with clerical texts seems unlikely.[33] Henry's recognition of the extent to which Nicholas's compositions were indebted to other authorities might have been the entire point.

Then again, one might be wary of where such an approach —showing evident sympathies with the New Philology movement— might lead. Lena Wahlgren-Smith, who recently published the long-awaited critical edition of Nicholas of Montiéramey's letters,[34] has quite rightfully expressed her concern regarding a "wholesale adoption" of the New Philological approach, which "assumes that all medieval literature, in all languages, all genres, and all periods, operates in the same way."[35] Such an approach is counterintuitive to those medieval attestations where value is attached to titled authority, where there is an outspoken preference for unviolated text, or where personal literary style is cultivated.[36] Bernard's denunciation of Nicholas for sending out texts without his consent serves as a firsthand example. Correspondingly, Nicholas's bold statement that Bernard's texts are in fact his own —"meo sensu inventos, meo stylo dictatos"—[37] also suggests that an explicit appropriation of texts by authors was not unknown in the twelfth century. From this perspective, Leclercq and Rochais had justifiable reasons to care about the interdependence of text and physical author (or performer). Constable has referred to a "rising tide of concern" over textual theft

[31] Moreover, the assertion that Bernard never heard of Nicholas again after he left Clairvaux is far from certain. See Constable, *The Letters of Peter the Venerable*, 2:330.

[32] "Here then is a case in which a skilled student of the *ars dictaminis* with alleged inclinations to forgery imitated a near-contemporary model, and we can assume that there would have been little difference between the 'honest' and dishonest imitation of Bernard's style," in Jaeger, "Prologue to the *Historia Calamitatum*," 13.

[33] Leclercq, "Sermons de Nicolas de Clairvaux," 57.

[34] Nicolaus Claraevallensis, *The Letter Collections of Nicholas of Clairvaux*, ed. and trans. Lena Wahlgren-Smith, Oxford Medieval Texts (Oxford, UK: Oxford University Press, 2018).

[35] Lena Wahlgren-Smith, "Editing a Medieval Text: The Case of Nicholas of Clairvaux," in *Challenging the Boundaries of Medieval History: The Legacy of Timothy Reuter*, ed. Patricia Skinner, Studies in the Early Middle Ages 22 (Turnhout: Brepols, 2009), 174.

[36] Minnis, *Medieval Theory of Authorship*, 10–1.

[37] Taken from MS Harley 3073, and edited in Leclercq, *Recueil d'études* 1:49–50.

in the late twelfth and thirteenth century, possibly instigated by rapidly changing approaches to "literary individuality."[38]

## 4.3 Bernard of Clairvaux's Letters

Bernard's epistolary corpus is very complex, but a coherent structure has been recognized thanks to Jean Leclercq's editorial achievements. Leclercq, whose terminology will be adopted here, has divided the corpus into a literary (*intra* corpus) and a non-literary section (*extra* corpus).[39] Bernard intended the letters in the *intra* corpus to circulate as a literary collection and kept refining them intensely throughout his life, whereas the second group of letters is scattered across time and manuscript traditions.[40] Then, within the first section, the literary or *intra* corpus, we can make another division. Manuscript transmission allows us to distinguish between letters written before Nicholas's arrival in the scriptorium and letters inserted later.[41] The letters that date from before 1145 in an earlier first appearance are found in the *brevis* manuscripts, whereas those added to Bernard's literary corpus afterwards can be found in the *perfectum* manuscripts. The *perfectum* corpus was assembled after Bernard's death in 1153, possibly by Geoffrey of Auxerre, and contains, aside from the *brevis* letters, many new additions (tables A.4.4 and A.4.5 give a detailed overview of which letters are included in either the *brevis* or *perfectum* samples).[42] It is important to note that those letters which were already found in the *brevis* manuscripts and reoccur in the *perfectum* corpus have sometimes been considerably worked over in the eight years between the two appearances. After all, Bernard's aim was to compose a unified piece of literature. He corrected, rearranged, and selected throughout his life. Importantly, Leclercq's edition of Bernard's literary letter corpus, which we use in these experiments, is almost entirely based on the *perfectum* transmission, which enjoyed the most popular circulation.[43] We therefore do not strictly work with the *brevis* corpus in its original pre-1145 form, but with a group of letters that was collectively reworked

[38] Constable, "Forgery and Plagiarism," 18 and 32.

[39] Leclercq separated the *intra* corpus (Epp. 1–310) from the *extra* corpus (Epp. 311–547) in the *SBO*.

[40] Leclercq, *SBO* 8:233–38.

[41] Leclercq, "Histoire ou littérature?," 158.

[42] The main difference between the *brevis* and *perfectum* manuscripts is that the latter contain versions of these letters that were clearly amended and lengthened. In the introduction to the edition of the *intra* corpus, Leclercq gives a full account of the arrangement of the different transmissions, in which he distinguishes three more or less homogeneous collections, two of which are the *brevis* and *perfectum* cycles, whose names have served as inspiration to how we labelled our chronologically ordered data. It should be noted that Leclercq mentions a third intermediary publication that we have decided to exclude from our main argument, namely the *longior* corpus, which was compiled by Geoffrey of Auxerre and was presumably published in 1145. The *longior* corpus already contains quite a few of the *perfectum* additions. However, this corpus is hard to date or reconstruct, making it less interesting for us to include in this study: see Leclercq, *SBO* 7:xv. We decided to make a distinction between the early *brevis* publication, when Nicholas of Montiéramey was certainly absent from Clairvaux, and the later publication, when both of them, or an even more developed chancery, could have exerted influence on Bernard's style.

[43] Leclercq, *SBO* 7:xvi.

and jointly disseminated. This condition of the texts is a flaw in the experiment that should be kept in mind during the analysis. Moreover, as mentioned earlier, Bernard's *extra* letters have not known a homogeneous transmission but have been handed down to us under divergent circumstances. The corpus therefore required some reorganization. Since Leclercq's edition allows us to assign individual letters to discrete periods, we decided to divide the *extra* corpus into three time-bound parts to see if Nicholas's arrival came with a stylistic impact: the first part dates from before 1140, the second between 1140 and 1145, and the third from 1145 onwards. Those *extra* letters that are of questionable dating and addressee have been left out of our experiments, for they cannot contribute to a study of Bernard's stylistic evolution through the influence of his secretaries (tables A.4.6–A.4.8 in the appendix I give a full overview of which *extra* letters were included or excluded).

Figs. 4.1 and 4.2 yield a respective PCA plot and (*k* nearest neighbors) network to calculate and visualize the stylistic differences between Bernard's letter corpus and Nicholas's authentic sermons and letters.[44] For each of these techniques, we have provided three additional subplots, which highlight how the different corpora are positioned within the clusters. There appear to be two general, observable dynamics, confirmed both by *k*-NN and PCA. Firstly, the writing style in Bernard of Clairvaux's letters is fairly coherent and forms a distinguishable cluster separated from Nicholas's works. Nevertheless —and this is the second, more hidden dynamic— our chronological rearrangements in the corpus have laid bare a gradual, subtle disturbance in Bernard's stylistic signal from 1140 onward, corresponding to the approximate time of Nicholas's arrival in Clairvaux,[45] and seemingly moving towards the latter's cluster. Yet, two major remarks are in order. Firstly, although the *perfectum* additions were indeed inserted into the literary corpus from 1140 onwards, some of them must have been first composed at a time before Nicholas's arrival. For example, sample 10 of the *perfectum* additions, which draws closest to Nicholas's cluster of all literary samples (only in the PCA, not in the *k*-NN network), contains letters that revolve around the schism between Antipope Anacletus and Pope Innocent II, a series of events that occurred between 1130 and 1138.[46] Although Nicholas was not yet part of Bernard's entourage during these events and was therefore likely not involved in their first redaction, he was nevertheless present when they were first sent out collectively with the other *perfectum* letters. Even if it might appear likely that if any refinement was imposed on these letters after 1140 Bernard, as the author, would have been most likely to do so, the latter's interference is technically plausible.

[44] See pp. 347ff. for a more detailed, technical explanation of PCA, and pp. 349ff. for network analysis. For the general use of PCA in computational stylistics, see Binongo and Smith, "The Application of Principal Components Analysis to Stylometry." A comprehensive introduction to network analysis can be found in Newman, *Networks. An Introduction*.

[45] That Nicholas's collaboration with Bernard started as early as 1140, when Nicholas carried the letters on Abelard's heresy to the Pope, is suggested by Turcan-Verkerk in Turcan-Verkerk, "*Ars Dictaminis*," 70.

[46] The sample contains Epp. 128, 134, 135, 137, 138, 140, 142, 144, 145, 146, 147 (*SBO* 7:321–51).

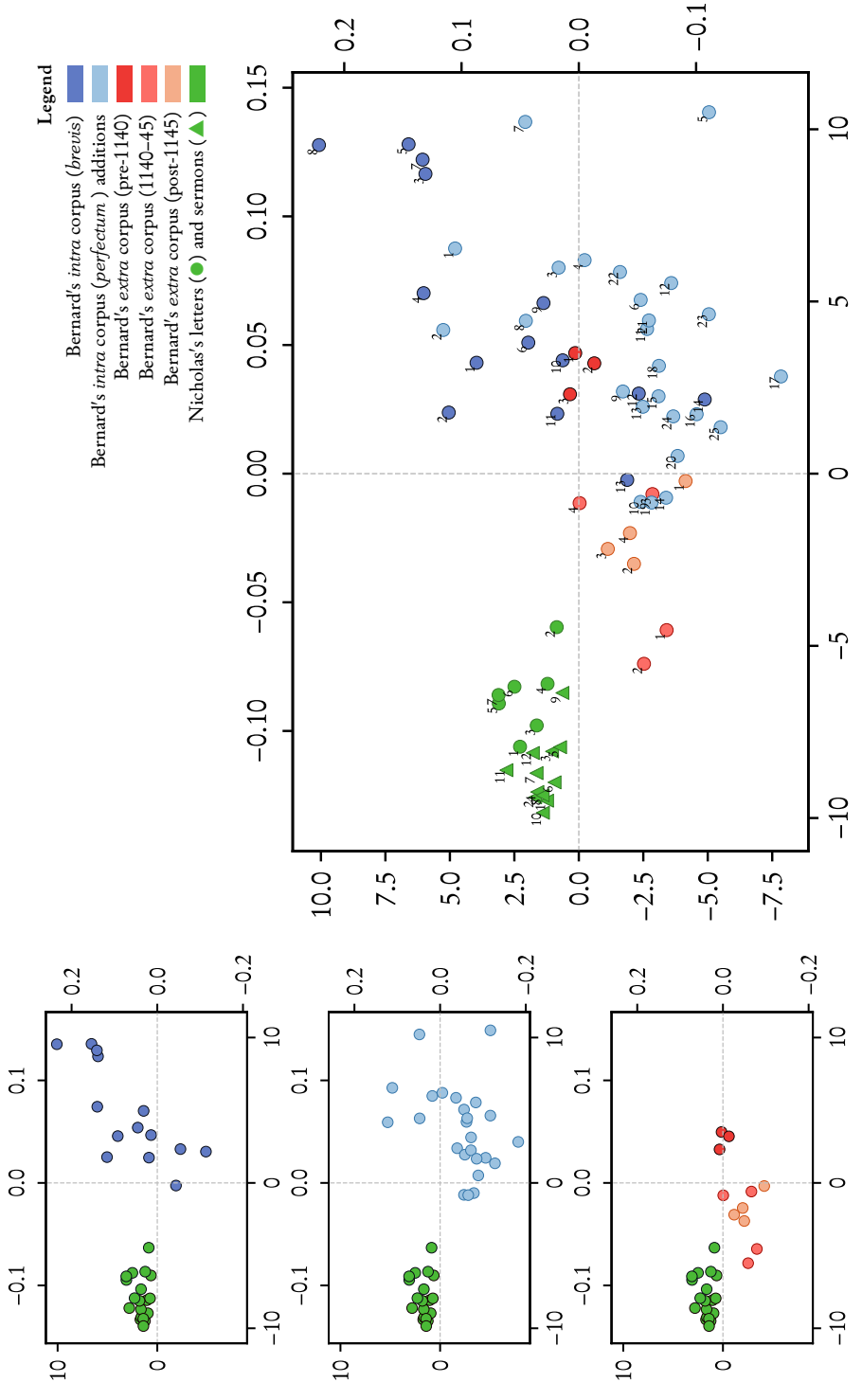



Figure 4.1: PCA plots of Bernard's epistolary corpus compared to Nicholas of Montiéramey's letters and sermons. The numbers on the right-hand plot indicate the sample index (corresponding texts found in the appendix on pp. 355). The left-hand subplots show the gradual, diachronic stylistic shift of Bernard's letters toward Nicholas as soon as he arrived in Clairvaux. Specific contents of the samples in this figure can be consulted in the appendix to this chapter (tables A.4.4–A.4.9). Settings: *s–l* = 3,000 | *type* = most-frequent function words | *n* = 150 | *vect.* = standard-scaled tfidf-weighted frequencies | *expl. var.* = 19.72%.

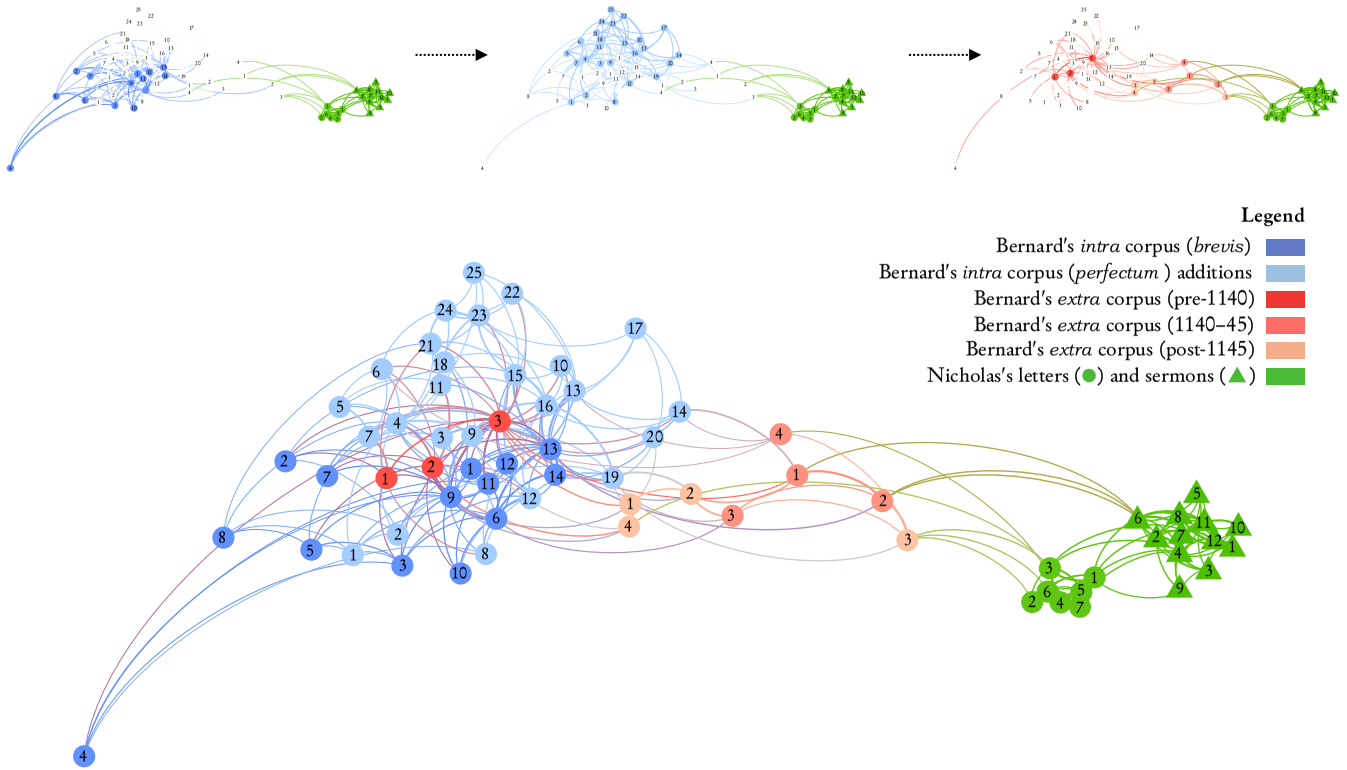



Figure 4.2: Network analysis of Bernard's epistolary corpus compared to Nicholas of Montiéramey's letters and sermons. The numbers indicate the sample index (corresponding texts found in the appendix on pp. 355)ff. The subplots in the top of the figure show the gradual, diachronic stylistic shift of Bernard's letters toward Nicholas as soon as he arrived in Clairvaux. Specific contents of the samples in this figure can be consulted in the appendix to this chapter (tables A.4.4–A.4.9). Settings: *s–l* = 3,000 | *type* = most-frequent function words | *n* = 150 | *vect.* = standard-scaled tfidf-weighted frequencies | distance metric = Minkowski.

The second remark ties in to the problem we have just raised. Although both figures show a diachronic stylistic shift, PCA slightly adjusts *k*-NN's inference that this shift has a determined direction towards Nicholas. The *extra* samples rather "float" around Nicholas's vicinity but never fully coincide. This suggests that the disturbances in Bernard's stylistic signal should not necessarily be as monocausal or directional as the *k*-NN network suggests, a nuance that reciprocates the historical skepsis raised in our first remark. Countless other variables aside from Nicholas's interference could have contributed to the subtle stylistic change in Bernard's letter corpus. One factor could be the lapse of time and Bernard's personal development. Another is the respective corpus's divergent transmission history. But perhaps the most crucial reason for PCA's less outspoken directionality is that Bernard did not have just one secretary. Although Nicholas was the scriptorium's headman, this experiment undoubtedly simplifies or fragmentizes its diversity of styles and personalities. We might even be surprised that Bernard's letters —considering the circumstances under which they were conceived— still display this amount of stylistic coherence (although there might have been a more outspoken divergence if we had been able to oppose the very original *brevis* corpus to the published versions).

This does not alter the fact that the plots' gravitation towards Nicholas's Latin style, which was of a very schooled nature, might hold some historical ground.[47] As Bernard became a public figure, he increasingly needed the support of scribes to take on administrative tasks, be it in Clairvaux or on exceptional occasions elsewhere.[48] These scribes

[47] Nicholas's style has often been deemed schooled and therefore unoriginal. See Constable, *The Letters of Peter the Venerable*, 2:328. Leclercq noted: "Ses exposés superficiels se développent selon un plan scolaire, en un style artificiel." See Leclercq, "Sermons de Nicolas de Clairvaux," 55. Likewise, Dorette Sabersky argued that "the syntactical structure of his sentences is similar to Bernard's, but often clumsier, less clear, less elegant, and rhythmically less balanced. His frequent use of word plays is at times rather superficial and, in opposition to Bernard's use, of little importance to the development of the contents. Repetitions of certain phrases and topics occur every so often. He favors rather unusual words and likes to quote classical authors. His literary exertions are only too obvious. All these aspects evidence Nicholas' lack of Bernard's creative spontaneity and mastery of language," in Dorette Sabersky, "The Style of Nicholas of Clairvaux's Letters", in *Erudition at God's Service*, ed. John Robert Sommerfeldt, Medieval Cistercian History 11, Cistercian Studies Series 98 (Kalamazoo: Cistercian Publications Inc., 1987), 196. However, it is all the more peculiar and contradictory to these former statements that —even very recently— attempts were made to attribute texts to Nicholas on the grounds of phrasing tics and certain lexical preferences (e.g. neologisms). See Patricia Stirnemann and Dominique Poirel, "Nicolas de Montiéramey, Jean de Salisbury et deux florilèges d'auteurs antiques," *Revue d'histoire des textes* 1 (2006): especially on 184. Either such attribution methods should be challenged (perhaps rightly so; their word and phrase concordances form particularly dangerous grounds for attributing authorship in a twelfth-century context that boasts such a high degree of 'plagiarism') or the statement that Nicholas has no style of his own should be withdrawn. I am convinced of the latter. Our computational experiments show that Nicholas, despite being Bernard's imitator, has a very controlled and rather clean authorial signal.

[48] Bernard would not necessarily have found help only in Clairvaux. It is conceivable that when he was occupied with the turbulent matters of the schism and was travelling through Italy he called for the assistance of papal scribes to whom he could dictate his messages. The papal notaries, educated in the *ars dictaminis*, would in fact have been schooled in a similar tradition as Nicholas, who had visited Rome and moreover corresponded with at least three popes during his lifetime. See Constable, "Dictators and Diplomats in the Eleventh and Twelfth Centuries: Medieval Epistolography and the Birth of Modern Bureaucracy," 43. This could also explain why *brevis* samples 12 and 13, which likewise have the schism as their subject, somewhat pair with letters that were added to the corpus later and not with the other *brevis* letters, which cling more closely together. The samples 12 and 13 contain Epp. 126, 127, 129, 130, 131, 132, 133, 136, 139, 141, and 143 (*SBO* 7:309–43). For a concise overview of Bernard's interference in the papal schism and its importance for his public

would have received a similar training or education. We can assume that most were under Nicholas's supervision, which meant that they departed from a common framework or set of rules from which they set out to imitate Bernard. This had become the nature of the epistolary writing art, or *ars dictaminis*, as introduced earlier in the first chapter of this thesis (pp. 27 ff.).[49] Letters were constructed on the basis of similar formulas, abounded in clever wordplay, and the rhythms of their prose pulsated under comparable cadences.[50] Diplomats, ambassadors, and secretaries would inspire one another in a network of correspondence or share these rhetorical devices within their scriptoria.[51] These practices might have considerably reshaped the stylistic homogeneity that is evident in the writings of Bernard from his earlier days, when he relied on a far smaller number of secretaries and had more time at his disposal so that he could be present during the various phases of composition. We know of Bernard's increasing discomfort concerning the fact that he felt obliged to delegate the writing of his letters and sermons to assistants, and his dissatisfaction with some of them when it came to grasping the *sensus* of his message.[52] Perhaps to his own frustration, Bernard was increasingly forced to have faith in the reliability of such scribes as Nicholas to reformulate his initial dictation in a letter that conformed to the style and content Bernard had intended. The *extra* letters would have received far less revision, resulting in the kind of hybrids that float towards middle ground in figs. 4.1 and 4.2.

career, see the subchapter "A Leading Figure in the Papal Schism 1130–38," in Brian Patrick McGuire, "Bernard's Life and Works: A Review," in *A Companion to Bernard of Clairvaux*, ed. Brian Patrick McGuire, Brill's Companions to the Christian Tradition 25 (Leiden: Brill, 2011), 40–7. For an elaborate read on the workings of the papal chancery, see Christopher Robert Cheney, *The Study of the Medieval Papal Chancery: The Second Edwards Lecture Delivered within the University of Glasgow on 7th December, 1964*, Edwards Lectures 2 (Glasgow: Jackson, 1966), 20–1: "Instructions about the framing of papal letters may be found in chancery ordinances and in guide-books for chancery clerks; these help to elucidate the legal principles which underlie the phraseology." The writing style of the papacy's chancery must have served as an important model to all clerks and diplomats both in ecclesiastical and worldly contexts.

[49] Witt, "Medieval 'Ars Dictaminis'."

[50] With 'comparable cadences,' I am here referring to rhetorical devices such as the *cursus*. See Janson, *Prose Rhythm in Medieval Latin*. The cursus has also been tested as a feature for authorship attribution. See Spinazzè, "'Cursus in Clausula'."

[51] Also see Constable, *Letters and Letter-Collections*, especially on 34–5: "This tendency towards a personalization of style and contents in eleventh- and twelfth-century epistolography was paralleled by a tendency, which was in some respects contradictory, towards formalization, which was represented by the emergence of the discipline known as the *dictamen* or *ars dictandi*, with teachers (*dictatores*), text-books (*artes* or *summae dictaminis*), and collections of model letters (formularies). Although *dictamen* now emerged for the first time as a discipline with clearly formulated rules, it had roots deep in the past and was connected in ways which are still not fully understood with the epistolographical rules and traditions which went back to Antiquity. [...] In the course of the twelfth century the number both of teachers and of text-books of *dictamen* spread rapidly, first in Italy and later, in the second half of the century, north of the Alps. Various schools developed with different styles, as at Bologna and Orleans; and although in the earlier twelfth century a certain number of writers, like St. Bernard and Peter the Venerable, who knew about *dictamen*, did not observe its rules, its influence was all but universal by the end of the century."

[52] This is evident in a frequently quoted letter of Bernard (Ep. 187), which is addressed to Peter the Venerable. The quote is given in full and discussed more extensively on p. 33.

## 4.4 Bernard of Clairvaux's Sermons

In this second visualization we put Nicholas's word to the test. Firstly, assuming that the secretary speaks the 'truth'[53] in his letter to Henry the Liberal (p. 128), which Leclercq cited as the most striking example of his plagiarism, we expect that a small number of sermons that occur in Bernard's *De diversis*, namely 6, 7, 21, 62, 83, 100, and 104, could be attributed to him instead. On the side, we test if his claims to Hugh's commentaries on the Psalms, which he also includes in the letter, hold any ground.[54] In a second phase, we follow up on Henri Rochais' conclusions that Bernard —not Nicholas— wrote *De diversis* 40, 41, and 42.[55] In fact, the *De diversis* collection in its entirety is worth testing here, as it suffers from some considerable issues of authenticity, provenance, and dating and might contain other traces of Nicholas's presence. The corpus comprises an assembly of unpolished and rudimentary sermons found in various, heterogeneous manuscripts, conceivably written down by secretaries and granted little revision by Bernard (unless if they were reused elsewhere).[56] Bernard never disseminated the *De diversis* sermons himself. They were gathered after his death and passed on for several centuries until Mabillon enumerated and published them in the seventeenth century. Leclercq and Rochais maintained Mabillon's structure in their edition. Secondly, we have included the *Sermones super Cantica Canticorum*, Bernard's literary masterpiece, as the cleanest possible specimen of Bernard's literary style to benchmark against these texts.[57]

Figs. 4.3 and 4.4 feature the results of matching up these texts. Firstly, when examining the visualizations, it is striking how the diversity of Bernard's *De diversis* is indeed captured. PCA, especially, demonstrates a discernible stylistic incoherence, as the samples burst open all over the plot (especially along the vertical axis of the second principal component), at times suggesting the interference of writers other than Nicholas or Bernard in their composition. Other samples gravitate in between Nicholas and Bernard, and in some cases Nicholas's influence on the style is undeniable. Before discussing some contingent subjects of interest, let us focus on the primary questions at hand. *De diversis* 6, 7, 21, 62, 83, 100, and 104, which Nicholas included in the letter to Count Henry the Liberal, do not betray an obvious affinity

[53] "The medieval idea of truth [...] was subjective and personal rather than, as today, objective and impersonal," in Constable, "Forgery and Plagiarism."

[54] Nicholas uses the phrase "aliosque sermones"in the prefatory letter in MS Harley 3073. See Leclercq, "Sermons de Nicolas de Clairvaux," 50. He thereby refers to the aforementioned sermons, a few other texts by Bernard, and, finally, Hugh of Saint-Victor's chapters on the Psalms gathered in the second book of his *Miscellanea* (*PL* 177:589).

[55] Rochais, "Sermons 40, 41 et 42."

[56] See Leclercq's introduction to the sermons, in Leclercq, *SBO* 6/1:59–71.

[57] Bernard must have started composing its beginnings around the end of 1135, but never commentated the entire Song of Songs. They are, nevertheless, regarded as his life's work and greatest literary achievement. See Leclercq, *SBO* 1:xv–xvi. Leclercq argued Bernard had passed away before he had the chance to finish his work, but it is more likely that Bernard never had the intention of discussing all the Canticles and has delivered us a finished work of literature. See Wim Verbaal, "Les sermons sur le cantique de saint Bernard: Un chef d'oeuvre achevé?," *Collectanea cisterciensia* 61 (1999): 167–85.

to Nicholas's style (although sample 1 is not far off). Neither are they unambiguously Bernard's. Both samples diverge strongly from Bernard's cluster and seem too hybrid in nature to be restrained to either of the authors' clusters. The case rather demonstrates how difficult it is to defend such concepts as 'single authorship' and 'textual theft' in a medieval context: the so-called [le] samples are clearly not of a 'singular' style (neither Nicholas's style nor Bernard's) but defy classification. In fact, if we compare both *k*-NN and PCA, Nicholas's influence in sample [le] 1 seems considerably larger than Bernard's. It has by now become an untenable simplification to argue that Nicholas has stolen these sermons, especially if we review the results of our second case, that of *De diversis* 40, 41, and 42 (four red samples labelled with [ro] of Rochais): although the sermons emanate from Bernardian thought, *k*-NN and PCA unambiguously cluster all three sermons together with those written by Nicholas, not Bernard.

There are some less straightforward developments on the side. Hugh of Saint-Victor's presence in both attribution problems remains somewhat unclear. Nicholas included Hugh's commentaries on the Psalms in his collection, yet figs. 4.3 and 4.4 show that he was unlikely to have been the (only) author of this incohesive text (see the purple samples, of which sample 9 comes closest to Nicholas).[58] Vice versa, *De diversis* 40 (which corresponds to the first half of the dubious [ro] samples) is collected in Hugh of Saint-Victor's *Miscellanea*. Would Nicholas have known Hugh well, and would they have collaborated before the latter's death in 1141? There is no proof of a direct acquaintance. Nicholas's musical sequences seem largely based on those of Adam of Saint-Victor († 1146), Hugh's choirmaster, but these texts enjoyed a popular circulation, so the similarity does not necessarily presuppose a personal tie.[59] For Bernard and Hugh, however, the connections are less far-fetched. We know they corresponded.[60] Hugh incorporated an entire letter he received from Bernard in his acclaimed masterpiece, *De sacramentis*.[61] Likewise, figs. 4.3 and 4.4 show that sam-

[58] Manuscript studies have argued that they can only be of Hugh's hand. See Joseph de Ghellinck, "Hugues de Saint-Victor," in *Dictionnaire de théologie catholique*, ed. Alfred Vacant and Eugène Mangenot, vol. 7 (Paris: Librairie Letouzey et Ané, 1922), 245. Although he admits that the *Miscellanea* is a confluence of the apocryphal and the authentic, de Ghellinck based his findings on the *Indiculum* of Hugh's writings. The commentaries on the Psalms often occur among Hugh's authentic works in the manuscript transmission. This has been confirmed in the exhaustive study of the dissemination of Hugh's oeuvre in Rudolf Goy, *Die Überlieferung der Werke Hugos von St. Viktor: Ein Beitrag zur Kommunikationsgeschichte des Mittelalters*, Monographien zur Geschichte des Mittelalters 14 (Stuttgart: Hiersemann, 1976), 58–63.

[59] "Since Nicolas is known for his plagiarism and incorporated the work of Hugh of Saint-Victor in the collection of his own opera dedicated to Count Henry, the suspicion arises that Nicolas modeled his work directly on that of Hugh's colleague, Adam of St. Victor," see John F. Benton, "Nicolas of Clairvaux and the Twelfth-Century Sequence with Special Reference to Adam of St. Victor," *Traditio* 18 (1962): 154.

[60] See for instance Ep. 77 under title of "Ad magistrum Hugonem de Sancto Victore,"*SBO* 7:184–200; also see Hugh Feiss, "*Bernardus Scholasticus*: The Correspondence of Bernard of Clairvaux and Hugh of Saint Victor on Baptism," in *Bernardus Magister: Papers Presented at the Nonacentenary Celebration of the Birth of Saint Bernard of Clairvaux, Kalamazoo, Michigan; sponsored by the Institute of Cistercian Studies, Western Michigan Univ., 10-13 May 1990*, Cistercian Studies Series 135 (Spencer: Cistercian Publications, 1992), 349–78.

[61] "Adding to the complications of *De sacramentis* as a text is Hugh's incorporation of passages not only from his

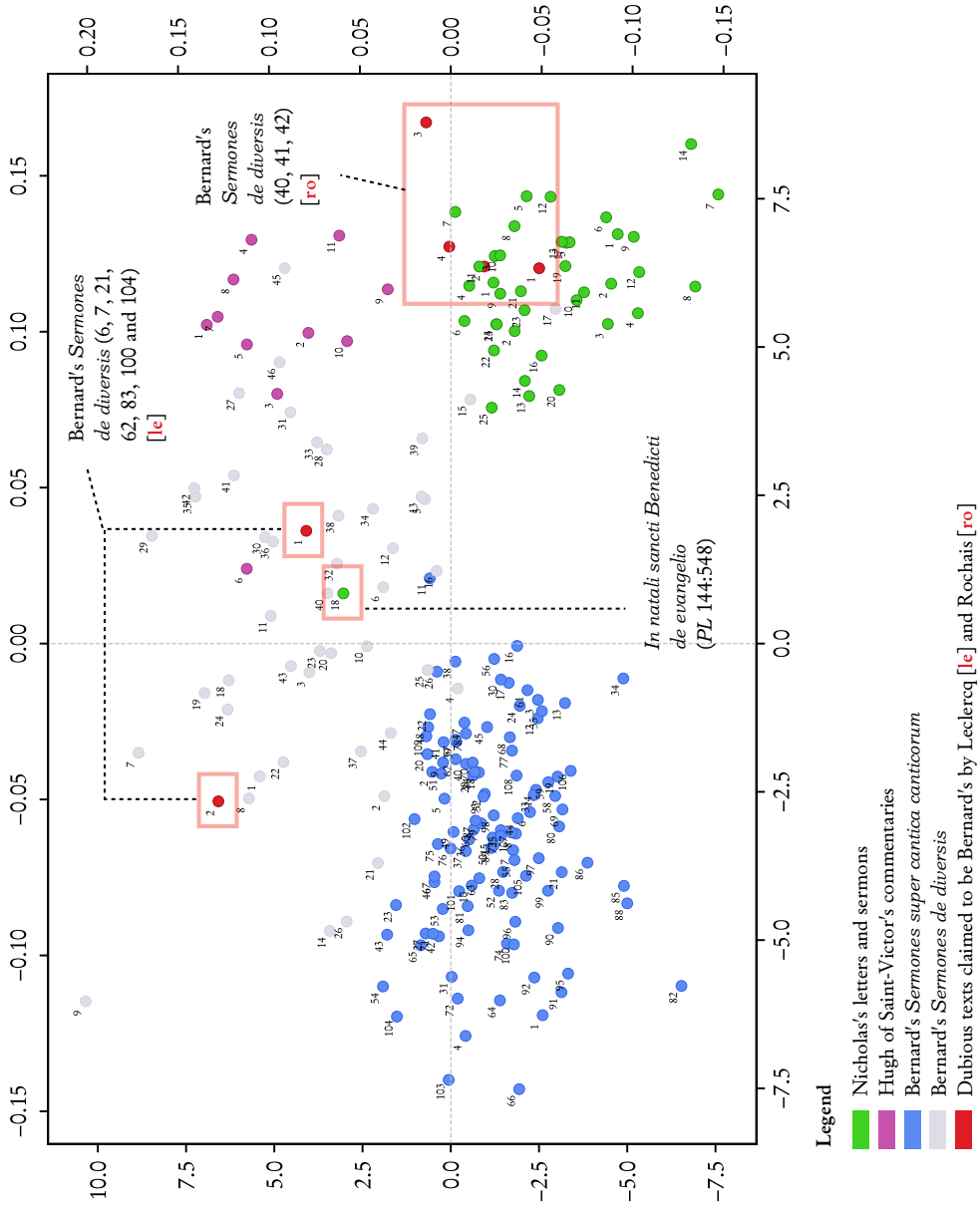



Figure 4.3: PCA plot of Bernard of Clairvaux's *Sermones super Cantica Canticorum* and *Sermones de diversis* compared to Nicholas of Montiéramey's letters and sermons and Hugh of Saint-Victor's commentaries on the Psalms. Specific contents of the samples in this figure can be consulted in the appendix to this chapter (tables A.4.10–A.4.12). Settings: *s–l* = 1,500 | *type* = most-frequent function words | *n* = 150 | *vect.* = standard-scaled tfidf-weighted frequencies | *expl. var.* = 11.68%.

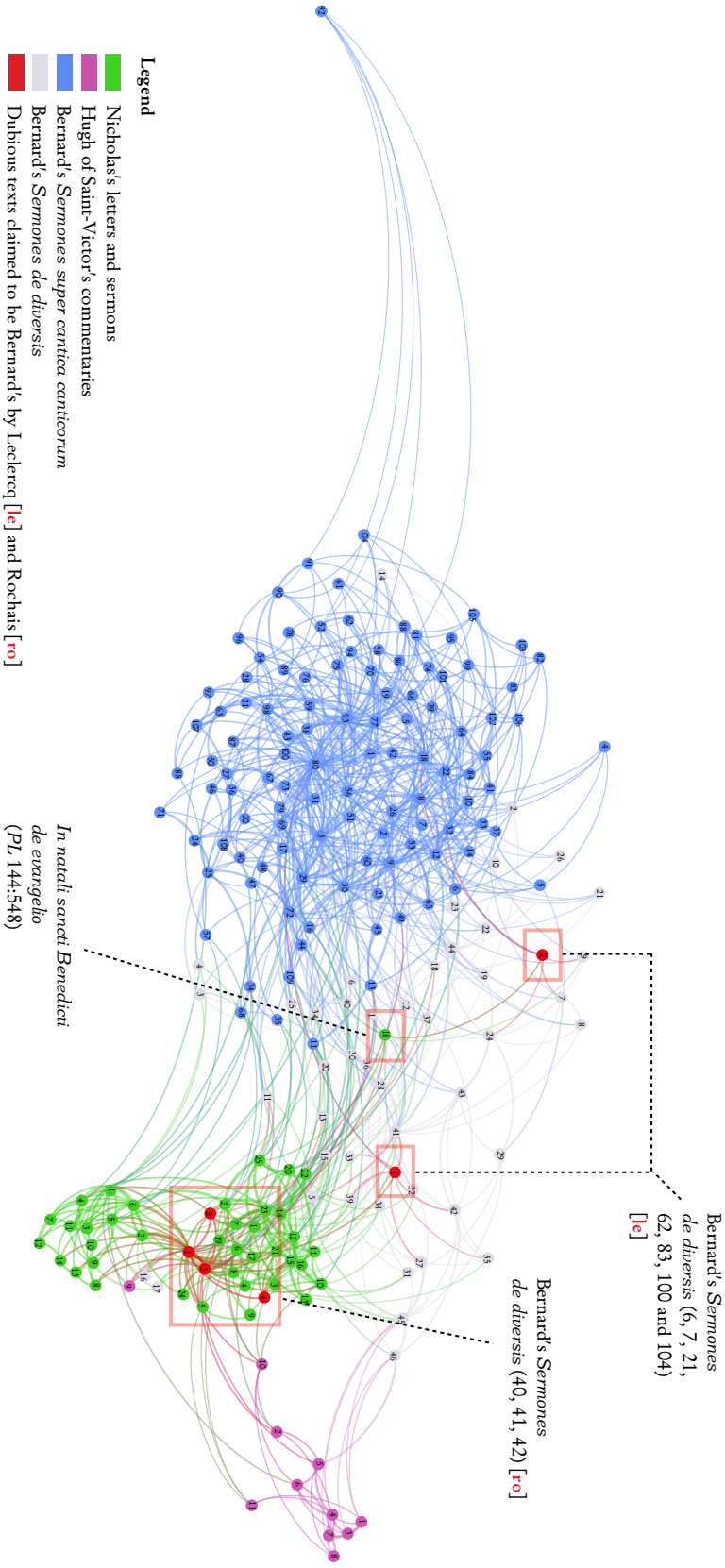



Figure 4.4: Network visualization of Bernard of Clairvaux's *Sermones super Cantica Canticorum* and *Sermones de diversis* compared to Nicholas of Montiéramey's letters and sermons and Hugh of Saint-Victor's commentaries on the Psalms. Specific contents of the samples in this figure can be consulted in the appendix to this chapter (tables A.4.10–A.4.12). Settings: *s–l* = 1,500 | *type* = most-frequent function words | *n* = 150 | *vect.* = standard-scaled tfidf-weighted frequencies | distance metric = Minkowski.

ples 46 and 47 of Bernard's *De diversis* bear some affinity with Hugh's commentaries. These samples comprise the very last additions to Bernard's corpus, *De diversis* sermons 112–25. They are shorter texts, which have not always been accompanied by the preceding sermons but must have circulated as a separate unit in manuscript transmission. Mabillon has argued that their provenance differs from that of the other *De diversis* sermons in a footnote,[62] thereby perhaps showing some wariness as to the authenticity of the works.[63] Although they might be Hugh's, we find that the textual style of both Bernard's *De diversis* and Hugh's commentaries is too unreliable to provide closure. The case for the triangular writing relationship between these authors is compelling, but there is insufficient historical proof to corroborate speculations of a collaboration between Nicholas and Hugh.

### 4.4.1 SVM Classification

As was emphasized in the previous chapter (from p. 96 onward), inspecting visualizations such as PCA and network graphs are informative, but one must remain wary of the fact that such figures drastically simplify high-dimensional data. Many additional informative dimensions in the data go lost. For the sake of further verification of figs. 4.3–4.4, a series of SVM classifiers (support vector machines) with different settings were optimized to segregate Nicholas's letters and sermons from Bernard's *Sermones super Cantica Canticorum*.[64] In optimizing the classifier, the following parameters were searched:

- **Feature type**: most frequent words (MFW), function words and 4-grams.
- **Weighting and scaling**: normal standard-scaled frequencies vs. tfidf-weighted frequencies.
- **Number of features**: 50, 100, 175 and 250 features.
- **C values**: 1, 10, 100 and 1,000.
- **Cross validation method**: stratified folds (taking into account class balance of training data) vs. 'normal' K folds.

own prior works but also from other theologians, patristic and contemporary, sometimes named but often without any attribution at all. In this respect, Hugh nicely represents the overall concern of twelfth-century authors to synthesize their sources," see Paul Rorem, *Hugh of Saint Victor*, Great Medieval Thinkers (New York: Oxford University Press, 2009), 60.

[62] Bernardus Claraevallensis, *Sämtliche Werke*, ed. Gerhard B. Winkler (Innsbruck: Tyrolia-Verlag, 1990–9), 9:882$^{Lh}$, 884$^{To}$. Mabillon relied on Jacobus Pamelius's (1536–87) edition of these sermons; see the note to *PL* 183:739.

[63] Mabillon was aware of the fact that Herwagen's and Pamelius's editions were to be approached with great caution when it comes to attribution. Herwagen's and Pamelius's collections of Bede's works are examples of how these editors "ignored and altered rubrics, expurgated passages, disregarded section breaks, and lied outright about the Bedan origins of their material," see Nathan J. Ristuccia, "The Herwagen Preacher and his Homiliary," *Sacris Erudiri* 52 (2013): 188.

[64] For a more detailed explanation of support vector machines (SVM), see chap. 3 on pp. 94ff., and this thesis's appendix on pp. 336ff. For its application to computational stylistics specifically, see Diederich et al., "Authorship Attribution with Support Vector Machines."

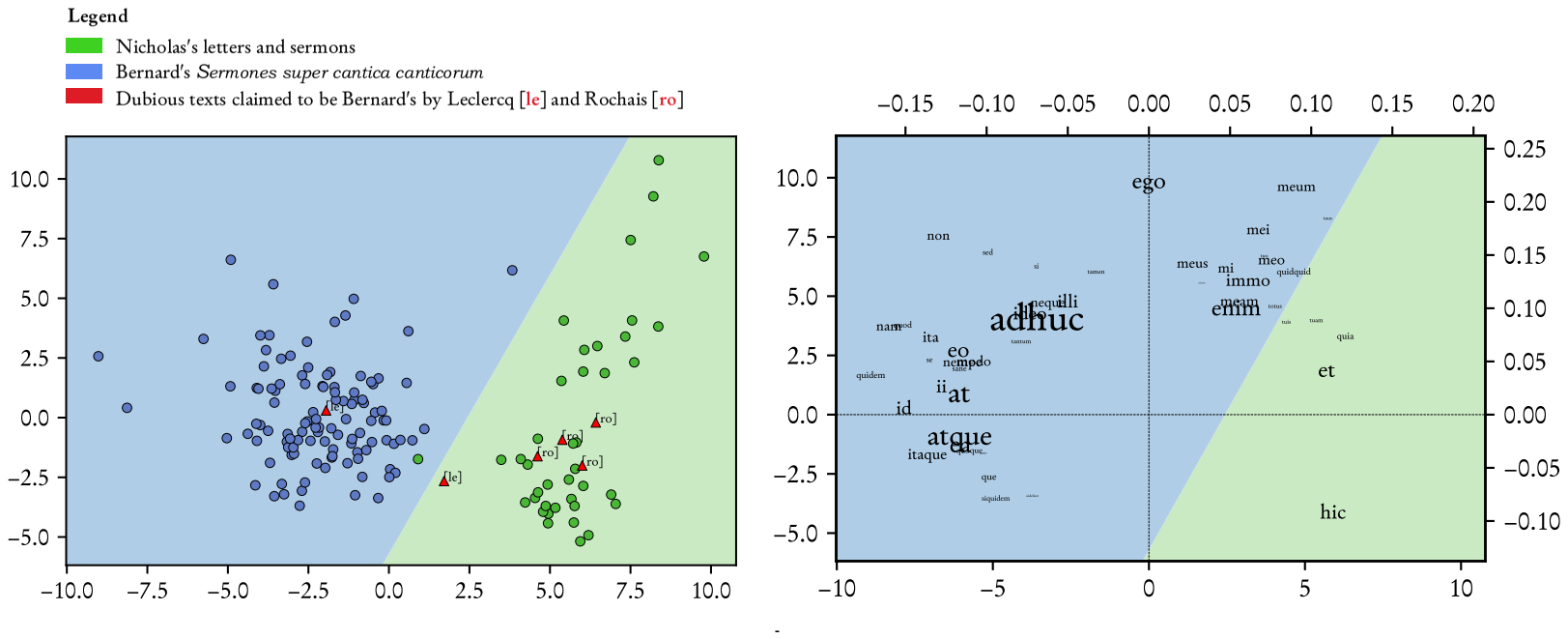



Figure 4.5: PCA plot giving a 2D intuition of the decision boundaries drawn by the best-scoring SVM classifier (whose settings can be consulted in table 4.1). The background colours indicate the respective 'zones' of candidate authors Bernard (blue) and Nicholas (green). The right-hand figure shows the 50 (out of 250) loadings. Only the highest ranking loadings on both PC's have been visualized (the font size expresses the weight). Settings: *s–l* = 1,500 | *type* = most-frequent function words | *n* = 250 (50 most important are visualized) | *vect.* = standard-scaled raw frequencies | *expl. var.* = 9.27%.

| Rank | Weighting | Feature type | Number | CV-method | Score |
|---|---|---|---|---|---|
| 1 | normal | function words | 250 | stratified | 0.992304 |
| 2 | normal | function words | 175 | stratified | 0.992304 |
| 3 | tfidf | function words | 50 | stratified | 0.992304 |
| 4 | normal | MFW | 250 | stratified | 0.992304 |
| 5 | normal | MFW | 175 | stratified | 0.992304 |
| ... | | | | | |
| 44 | normal | MFW | 50 | KFold | 0.700194 |
| 45 | normal | 4grams | 175 | KFold | 0.699900 |
| 46 | tfidf | MFW | 50 | KFold | 0.697660 |
| 47 | normal | 4grams | 50 | KFold | 0.685919 |
| 48 | tfidf | 4grams | 50 | KFold | 0.683534 |

Table 4.1: Ranked results of SVM models' performance in distinguishing between the works of Bernard of Clairvaux and Nicholas of Montiéramey. There were 192 models in total, 48 of which were retained due $c$=1 consistently being the highest performing value. The larger mid-section is not shown, which allows to spot the parameters causing the main differences between high and low performance.

In total, the training process yielded some 192 SVM models. Only 48 of these were retained, as the $c$ parameter proved consistently optimal at $c$=1. These last 48 classifiers exhibited a wide diversity in average performance, ranging from an average of 68.35% to 99.23%.[65] Table 4.1 yields a ranking of the average performances for the top 5 and lower 5 SVM models in distinguishing between Bernard's and Nicholas's different writing styles. From this table it appears that stratifying the training data is beneficial (as opposed to an equal folding technique, KFold), and that more features works favourable as opposed to fewer features. Perhaps surprisingly, the scaling and weighting techniques and the feature types made relatively fewer difference, even though 4-grams consistently performed worse than lexical items (MFW and function words).

Consequently, from all of these classifiers, one may choose the highest-ranking one. The PCA plots in fig. 4.5 give a 2D intuition of the highest-ranking classifier's success in segregating the two authors' works (99.23%). As one can see, the coloured backgrounds, separated by decision boundaries, indicate Nicholas and Bernard's respective stylistic 'zones' (green and blue) as they were demarcated during training.[66] The disputed [le] and [ro] sermons (in red), our test data, are plotted over the respective zones, thereby choosing sides between either Bernard or Nicholas. It turns out that five out of six sermons which Rochais and Leclercq argued to be exclusively Bernard's, are heavily indebted to Nicholas's style.

Instead of solely relying on the result of the best classifier, ranking first in table 4.1 and visualized in fig. 4.1, one could also determine which prediction the majority of

[65] These percentages were based on a mean over accuracy, precision and recall, the concepts of which are explained on p. 342.

[66] Note that the original classifier's decision boundary —which cannot be visualized here— takes into account many additional dimensions of information in its decision function.

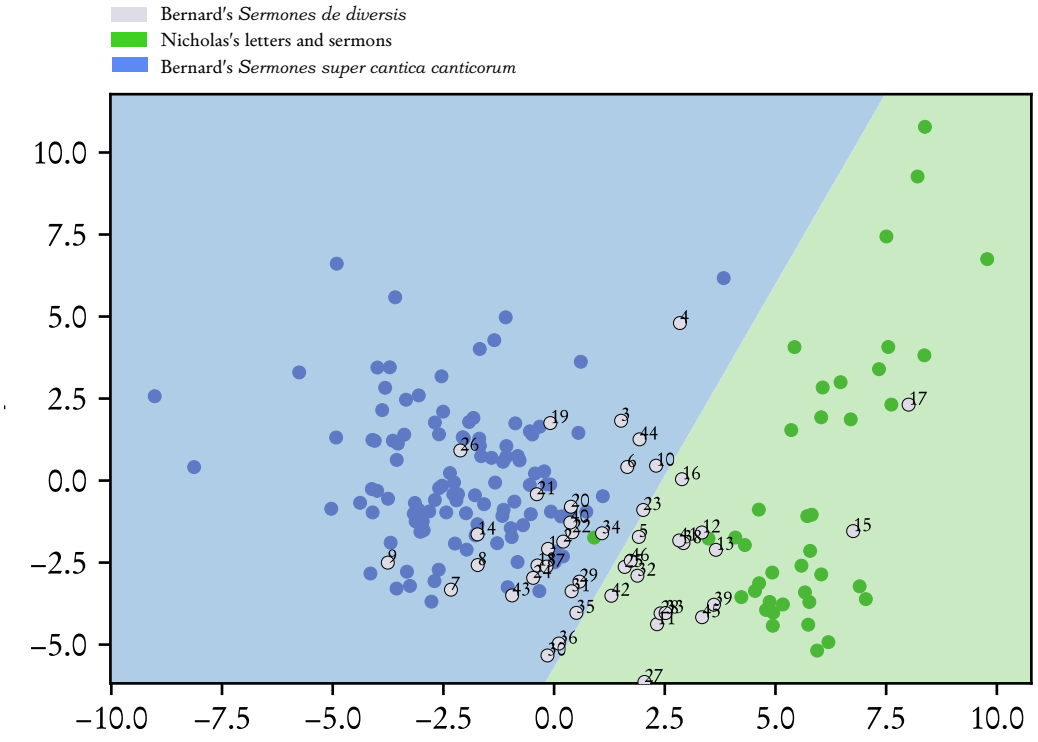



Figure 4.6: PCA plot giving a 2D intuition of the decision boundaries drawn by the best-scoring SVM classifier (whose settings can be consulted in table 4.1). The background colours indicate the respective 'zones' of candidate authors Bernard (blue) and Nicholas (green). The full *Sermones de diversis* corpus is plotted on top of the decision boundary, indicating their divided loyalties to both authors. For contents of the samples, see the addenda to this chapter, more specifically table A.4.10 on p. 358. Loadings can be inspected on the right-hand side in fig. 4.5. Settings: *s–l* = 1,500 | *type* = most-frequent function words | *n* = 250 (50 most important are visualized) | *vect.* = standard-scaled raw frequencies | *expl. var.* = 9.27%.

classifiers made. In doing so, one takes into account some inferior models which relied on inferior settings and other features subsets (such as MFW or 4-grams), but these might in reality yield an honest result of the extent by which a particular sample's attribution lingered 'in a grey zone.' By chaining up a series of attributions made by different classifiers we gain a naive type of confidence score. The results of this are tabularized in table 4.2, where the prediction is indicated in the middle column, and the confidence score (majority) in the right-hand column. Here, for instance, we learn that for sample [le]_2, already a doubtful case in the PCA plot and *k*-NN network (figs. 4.3–4.4), some disagreement amongst the different classifiers exists. The attribution to Nicholas is 88% secure, meaning that 12% of the classifiers still prefer Bernard as the better candidate. This confidence score allows to think of Nicholas and Bernard's collaboration in degrees, instead of in strict, binary categories.

The PCA plot in fig. 4.6 may give further illustration of this idea of 'distributional' rather than individual authorship. If one plots additional text samples from Bernard's *Sermones de diversis* on top of these decision boundaries, one sees many more affini-

| Sample | Prediction | Confidence |
|---|---|---|
| [ro]_1 | Nicholas | 1.0 |
| [ro]_2 | Nicholas | 1.0 |
| [ro]_3 | Nicholas | 0.96 |
| [ro]_4 | Nicholas | 0.92 |
| [le]_1 | Bernard | 1.0 |
| [le]_2 | Nicholas | 0.88 |

Table 4.2: Attribution results of disputed sermons. The prediction made by the majority of 48 classifiers is indicated in the middle column. The right-hand column indicates a naive kind of confidence score.

ties to Nicholas's profile appear. This gives yet another indication of how the corpus of Bernard of Clairvaux, as it is transmitted to us, is a collaborative product, and small artefacts that have survived in the text may serve to map out the dynamics at play in the Clairvaux scriptorium.

## 4.5 Conclusive Remarks

Jean Leclercq, in aspiring to discern the psychological personality of the author behind any given historical text, conceded the difficulty of infiltrating the "screen of rhetoric" so characteristic to twelfth-century literature, referring to its predilections of imitation and formal rigidness.[67] The surface of the medieval text can strike one as impenetrable. In a similar vein, Giles Constable has argued for medieval epistolography —and he may well have found the statement applicable to all twelfth-century texts— that "style alone is not a reliable guide to authorship," and that "even today some of the works of Nicholas of Montieramey, who was clearly an accomplished mimic, are not easy to distinguish from those of Bernard and other writers."[68] Yet this trait of medieval texts, which is primarily qualitative and open to subjective interpretation, is elusive only in a close-reading approach and seems not to present a problem when form is quantified in a distant-reading approach.[69] Computational stylistics disables the distracting semantics in which Nicholas's style is embedded, and patterns the turns of phrase that reveal his presence (or that of a chancery working under his lead). It can only follow, then, that Nicholas's reputation of being Bernard's pale shadow is a construction by readers who have undoubtedly experienced the difficulty of peering through the curtain of imitation, citation, and formalization when it comes to recognizing the author behind the text. It turns out that, if Nicholas's style is not 'distinguished,' in the sense that it can be judged as of a high literary value, it is nonetheless distinguishable. This does not simply mean that the application of computational stylistics results

[67] Leclercq, "Modern Psychology," 476.

[68] Constable, *Letters and Letter-Collections*, 50.

[69] In this regard one may think again of the prejudiced accusations at Nicholas's address that the latter has no style of his own. See n. 47.

merely in giving an individualized coloration to the question of authorship. A glance at each of the figures in this article demonstrates the interconnectedness (or "infinite shadings,"[70] in Constable's words) laid out as networks between these two authors. Computational stylistics therefore does not simply force us to choose a side in the medieval authorship dilemma, which is infinitely fought out along the axes of the 'individual' and the 'distributional.' It rather becomes these axes and reenacts the tension field as is. Neither is Nicholas's and Bernard's collaboration depicted as a hierarchical author-scribe relationship in one-sided text classifications, nor must we seek refuge in a stopgap conception of infinite authority and authorship. This approach embraces both an acknowledgement that the practice of cooperative medieval authorship is complex, and a refusal to believe that medieval authorship is interminably diffuse. Therefore, computational stylistics provides valuable tools with which to validate or contradict contrasting theories with objective material, taking the voices from the past at face value and opening up avenues to rethink our approach to medieval texts in literary theory, text editing, and historical studies.

[70] Constable, "Forgery and Plagiarism," 3.

# 5

# Siblings' Synergy: Elisabeth and Ekbert of Schönau*

## 5.1 Elisabeth and Ekbert

### 5.1.1 Context

Identifying the author behind the Benedictine nun Elisabeth of Schönau (1129–1164/5) has proven a difficult balancing act. We know very little of the German visionary, and her writings present too much of an ambiguous source to gain a reliable acquaintance. In attempting to describe her literary production, most scholars have been forced to understand and access the figure of Elisabeth through others. These 'others' are figures to whom she can be compared, or with whom she is closely related. Today still, Elisabeth is seldom discussed without reference to Hildegard of Bingen (1098–1179), who is subject of the next chapter. Hildegard, who was thirty-one years older and likewise provenant from the Rhineland, was Elisabeth's predecessor and main source of inspiration. Two of Elisabeth's letters addressed to Hildegard (in a collection consisting of twenty-two letters in total) and one response letter testify that the two corresponded.[1] Also in the transmission of their texts, the Rhenish nuns were occasionally paired up.[2] Despite some suggestions —based on the circulation of manuscripts— that Elisabeth may have been more popular than Hildegard from

* My gratitude goes out to Caroline Van Sumere, who during her time as a research intern provided me with the prospective research that lies at the basis of this chapter.

[1] The two letters of Elisabeth to Hildegard (numbered 18–19), were integrated in Elisabeth's *Liber tertius visionum*, see Elisabeth Schoenaugiensis, *Visionen*, 70–4 (chapters 19—21). An alternative version of letter 18, originally edited by the Parisian Jacques Lefèvre d'Étaples (Jacobus Faber Stapulensis) in 1513, is found under *PL* 195:120–4. Hildegard's response is also found in Roth's edition on 178–9.

[2] A well-known example is the Dendermonde codex which contains Hildegard's *Liber vitae meritorum*, followed by Elisabeth's *Liber viarum dei*. The manuscript was originally named MS Dendermonde, Abbey of Saint Peter and Paul, 9, but has since 2017 become the possession of the Maurits Sabbe Library of the Catholic University of Leuven. For Elisabeth's inclusion in the codex, see Kurt Köster, "Das visionäre Werk Elisabeths von Schönau. Studien zu Entstehung, Überlieferung und Wirkung in der mitteralterlichen Welt," 4 (1952): 114.

the Middle Ages up to even Early Modern times, she definitely became Hildegard's epigone in the centuries after.[3] In a similar vein, Elisabeth is often discussed in tandem with her brother, abbot Ekbert of Schönau († 1184). Ekbert was the spiritual guide and confessor to the cloister community, and Elisabeth's writing collaborator. When Elisabeth reached the age of twenty-five, some three years after the transcription of her first visionary text to parchment, Ekbert would come to join her in Schönau (1155) to start up a literary collaboration that would last until the end of Elisabeth's life (ten years later). The number of tasks that consequently fell under his responsibility —or which he appropriated for himself— were various: editor, transcriber, inspirer, interrogator, publisher, censor... Judging by these various and important responsibilities, Ekbert took on a significant role in his sister's literary production. Before his conversion to monasticism, Ekbert had been canon in the Cassius church of Bonn. At the surprise of some of his colleagues,[4] Ekbert traded in his career ambitions in the Church for a monastic life in the double house of Schönau.[5]

Elisabeth's literary influences and her constant collaboration with her brother presents one of the most symptomatic and difficult cases of collaborative authorship in the twelfth century. According to Ekbert's testimony in the *Narratio Eckberti*, his short introductory passage opening Elisabeth's collected visions,[6] Elisabeth would have uttered her visions in Latin, in German, or in a mixture of both. No German records of her are existent. In a way reminiscent of Hildegard, it is described as miraculous that Elisabeth was capable of uttering Latin: "[...] suddenly she proclaimed some of the most divine words in the Latin tongue, which she had not ever been taught by anyone else, and which she could not have found out by herself since she was uninstructed and possessed little to no proficiency in Latin speech."[7] The assumption is that Elisabeth would have not known Latin, and may have never written a word herself, but all of

[3] Kurt Köster argued on the basis of Elisabeth's wide manuscript dissemination —be it still quite some time before critical scholarship on Hildegard took place— that Elisabeth might have been more popular than Hildegard. See Kurt Köster, "Elisabeth von Schönau: Werk und Wirkung im Spiegel der mittelalterlichen handschriftlichen Überlieferung," *Archiv für mittelrheinische Kirchengeschichte* 3 (1951): 243; although there are approximately 360 codices known worldwide for Hildegard of Bingen, Anne Clark seemed to consent to Köster's suggestion by stating that "the modern preference for Hildegard stands in contrast to the apparent medieval preference for Elisabeth's works, witnessed by the substantially greater number of manuscripts transmitting them." See Michael Embach, "Hildegard of Bingen (1098–1179): A History of Reception," in *A Companion to Hildegard of Bingen*, ed. Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, Brill's Companions to the Christian Tradition 45 (Leiden: Brill, 2014), 273; and Anne L. Clark, *Elisabeth of Schönau: A Twelfth-Century Visionary*, Middle Ages Series (Philadelphia: University of Pennsylvania Press, 1992), 5. For illustrative purposes, "there are twenty-nine known twelfth- and thirteenth-century manuscripts of the *Liber viarum dei* and eight of the complete Scivias." See ibid., 36.

[4] So much becomes clear from Emecho of Schönau's *Vita Eckeberti.* See Clark, "Repression or Collaboration?," 153, n. 12. The *Vita* is edited in S. Widmann, "Vita Eckeberti," *Neues Archiv* 11 (1886): 447–54.

[5] On the history of the Schönau double monastery, see Joachim Kemper, "Das benediktinische Doppelkloster Schönau und die Visionen Elisabeths von Schönau," *Archiv für mittelrheinische Kirchengeschichte* 54, no. 1 (2002): 55–102.

[6] From "Fuit in diebus [...]"to "[...] inicium erat huiusmodi." See Elisabeth Schoenaugiensis, *Visionen*, 1–2.

[7] "[...] subito verba quedam divinissima latino sermone proferebat, que neque per alium aliquando didicerat, neque per se ipsam adinvenire poterat, utpote inerudita et latine locutionis nullam vel minimam habens periciam," see ibid., 2.

this is questionable.[8] Once dictated orally, Elisabeth's visions were to be redacted on parchment. In the first years of her writing career this would have been a task undertaken by her fellow nuns. From 1155 onwards, it is assumed to have become her brother's.

The assistance which Elisabeth received in the redaction of her texts, both before as after Ekbert's arrival, has been a much debated question in attempting to better understand Elisabeth and the composition context of her texts. For instance, it is uncertain if Ekbert exerted influence on her texts only as soon his writing partnership had been institutionalized by the monastery's presiding abbot Hildelin, or if that was already the case before his arrival in 1155.[9] If the former is true, Elisabeth's visions dating from before Ekbert's arrival could perhaps testify more of her own voice than her later visions containing Ekbert's intrusions or "repressions."[10] The truth of the matter is, however, that we cannot possibly retrieve the oldest versions of the visionary's first texts. We can conjecture their original composition date by clues given in the text or on the basis of the calendar of saints,[11] but the earliest redaction of Elisabeth's texts circulating outside the walls of Schönau is found only in 1159,[12] which is still some four years after Ekbert's arrival, and seven years after when Elisabeth composed her earliest texts. In the time lying in between Elisabeth's first composition and her first 'publication' —if one allows the anachronism—, Ekbert had ample opportunity to leave a mark. The existent different redactions of Elisabeth's collected works teach us that he was prone to revise Elisabeth's works during the last years of her life, and even more intensively after her passing in 1164/5.[13]

The word 'repression' as I used it above implies that the interventions of Elisabeth's female assistants would have been more reticent than those of Ekbert's, which must remain questionable. If one can assume that Elisabeth's Latin was as rudimentary as she lets off in her visions —for which I share the reservations formerly expressed by such scholars as Clark—,[14] we have to assume that her female secretaries (were there any

[8] "It is not clear from the visionary records that Elisabeth had any part in the production of the texts other than narrating her experiences. While it has been suggested that Elisabeth apparently wrote down many things herself, the evidence for this is questionable," see Clark, *Elisabeth of Schönau*, 53.

[9] According to Clark, Ekbert's influence took place before his arrival date. See ibid., 32.

[10] Clark, "Repression or Collaboration?"

[11] Many of the visions set out by specifying the day according to its feast in the church calendar, and even the time is mentioned following the hours of the divine office. On the dating of composition, see Köster, "Das visionäre Werk Elisabeths von Schönau," 80.

[12] Clark, *Elisabeth of Schönau*, 45–6.

[13] On the different redactions of Elisabeth's works, see Köster, "Das visionäre Werk Elisabeths von Schönau," 83; and the summary and improvements of Köster's research provided in the appendix of Clark, *Elisabeth of Schönau*, 137–45.

[14] Clark, *Elisabeth of Schönau*, see especially from 30 onward: "Considering that Elisabeth probably came from a well-established family and that she entered the convent at age twelve, it seems quite likely that she could have been educated in Latin. Even though Ekbert wishes to diminish her Latin skills, nowhere in the entire corpus does he hint that Elisabeth had difficulty understanding Latin, and his description of Elisabeth as having 'little or no skill in Latin speech' evinces a certain hesitance to assert complete ignorance."

male ones aside from Ekbert?)[15] equally had to translate and revise Elisabeth's speech. It takes a long stretch of the imagination to assume that whilst being surrounded by educated female scribes from similar backgrounds, Elisabeth herself did not know sufficient Latin to compose her own works. It is probably wiser to interpret Elisabeth's assertions of inadequacy as a topical and strategical devaluation of her own capacities as a writer, reminiscent of those made by Hildegard of Bingen.[16]

Such considerations aside, what can nevertheless be concluded is that Ekbert's physical presence in Schönau kickstarted the public promulgation of Elisabeth's literary output.[17] It is also clear —as Clark and Dinzelbacher have stressed— that in the visions composed after Ekbert's arrival the latter's theological agenda occasionally shimmers through quite overtly in her texts. An often cited example is Elisabeth's developing interest in Catharism, featuring most prominently at the end of her third vision book. The movement had been vehemently protested against by Ekbert in sermons of his own,[18] although Anne Clark argued Elisabeth's integration of the Cathar movement to have been different "in style and message."[19] There have been many scholars who assert to have discerned Ekbert's stylistic interventions within fragments of Elisabeth's works.[20] Ekbert's interventions are said to appear most conspicuously "beginning at the nineteenth chapter of the second *Liber visionum*."[21] If this is true, Ekbert's assertions that he would have left Elisabeth's speech unchanged —as he does, for instance, in the prologue to Elisabeth's visionary collection (*Prologus Eckberti*)— becomes suspect.[22]

[15] As early as under its first abbot Hildelin († 1165/7), Schönau was a double monastery under influence of the Hirsau abbey in the northern black forest. The original buildings were burned down in 1783, but Kemper's study suggests that Schönau was made up of duplicated buildings and separate chapter houses for men and women, and by all likelihood the nuns used an oratory instead of the monastic church of Saint Florian. Obituaries of the twelfth- to fourteenth centuries testify of the male monks and abbots living in cohabitation with the nuns. See Kemper, "Das benediktinische Doppelkloster Schönau"; and Flanagan, *Hildegard of Bingen, 1098–1179: A Visionary Life*, see especially 37–8. Flanagan moreover states that "details from the works of Elisabeth of Schönau suggest that a more relaxed form of association was common in German double monasteries. We have, for example, a picture of Elisabeth leaning out of her window to talk to a monk and a description of a group of nuns standing in conversation around a priest on his way to bring the reserved elements to a sick nun. Elisabeth knows the names and recognizes the faces of certain members of the male community when she sees them in a vision. Such evidence suggests that the relationship between monks and nuns in a double monastery might be less than rigorously controlled"; the window passage can be found in the first book of visions, see Elisabeth Schoenaugiensis, *Visionen*, §20, 12: "Et venit unus ex fratribus ad fenestram, et rogavi eum, ut missam de sancta trinitate celebraret, et annuit," which translates to "One of the brothers came to the window and I asked him to celebrate the Mass of the Holy Trinity and he agreed"; translation taken from Elisabeth Schoenaugiensis, *Complete Works*, §20, 56.

[16] This point is made more elaborately in Elisabeth Gössmann, "Das Menschenbild der Hildegard von Bingen und Elisabeth von Schönau von dem Hintergrund der frühscholastischen Anthropologie," in *Frauenmystik im Mittelalter*, ed. Peter Dinzelbacher and Dieter R. Bauer (Stuttgart: Schwabenverlag, 1985), 41-2.

[17] Ekbert would only become abbot of Schönau after Elisabeth had passed. Clark, "Repression or Collaboration?," 154.

[18] The *Sermones contra catharos*. See *PL* 195:11–97.

[19] Clark, "Repression or Collaboration?," 161.

[20] Roth, *Visionen*, cix–cx; Clark, *Elisabeth of Schönau*, 63.

[21] Coakley, *Women, Men and Spiritual Power*, 28.

[22] Ekbert stated that during the redaction he would not add or subtract anything from Elisabeth's dictation: "Conscripsi omnia hec [...] ita quidem, ut ubi erant Latina verba angeli immutata relinquerem, ubi vero teutonica erant, in latinum transferrem, prout expressius potui, nihil mea presumptione adiungens, nihil terreni commodi querens, testis mihi est deus,

### 5.1.2 The Composition of Elisabeth's Visions

It is generally assumed that the redaction process of Elisabeth's visions was not a unilateral dictation on account of Elisabeth, but that Elisabeth's visions were responses to prompts given by Ekbert.[23] In all likelihood Ekbert himself was responsible for including his presence in Elisabeth's visions, a "self-inclusion [...] to show that he has asserted his *own* direction over Elisabeth unambiguously."[24] Aside from Ekbert's role as a kind of discussant, Clark has also brought forth evidence suggesting that Ekbert would only have disseminated parts of Elisabeth's writing which he considered appropriate.[25] In addition to this censorship, the opposite also seems to have been customary. Ideas which Elisabeth feared would go against contemporary orthodoxy, such as the bodily assumption of the virgin Mary in her *De Resurrectione*,[26] may have been published by Ekbert against her will. The text ultimately became one of her most popular writings.[27] Ekbert, then, can be thought of as a kind of gatekeeper, who exerted a considerable influence as to which aspects of the author Elisabeth were disclosed to contemporary and/or future readership. He moderated Elisabeth's output, and managed to what extent her horizons could be widened. Ekbert's authority was therefore not merely situated on a creative or artistic level, but can be extended to the role of artistic director or publisher *avant la lettre*.

In consideration of Ekbert's influence both materially as text-internally, different approaches have been proposed by which to recognize what is 'authentic' in Elisabeth's texts. Very often, 'authentic' in these contexts becomes synonymous to 'Elisabethan.' With the surge of feminist approaches in the humanities, the concern of uncovering the female voice has increasingly attracted attention. Most scholars agree that Elisabeth's voice is embedded in her visions in some way. But in which way? And how relevant is this question in light of interpreting the text? Interestingly, although the applicability of individual authorship in the Middle Ages has been questioned in the recent decades (as has been elaborately discussed in chapter 1), no serious form of scholarship can discuss Elisabeth of Schönau without sidetracking into the difficult question of her authorship. Too much depends on it. In order to speak of the author Elisabeth of Schönau, to speak of her thought, and of the context in which her literary production

cui nuda et aperta sunt omnia." Elisabeth Schoenaugiensis, *Visionen*, 1.

[23] Clark, *Elisabeth of Schönau*, 51.

[24] John Wayland Coakley, "A Shared Endeavour? Guibert of Gembloux on Hildegard of Bingen," in *Women, Men, and Spiritual Power: Female Saints and Their Male Collaborators*, ed. John Wayland Coakley (New York: Columbia University Press, 2006), 26.

[25] Clark, *Elisabeth of Schönau*, 53.

[26] See Clark, *Elisabeth of Schönau*, 40. "In this text, Elisabeth records her discomfort at announcing something that conflicted with tradition and her hesitation to publish something that would reveal her to be an *inventrix novitatum*. [...] Despite Elisabeth's intended limitation of the audience for these revelations, this text was included in all the major versions of the visionary collections, a fact which suggests that Ekbert did not always heed Elisabeth's own concerns about the publication of her visions"; the short vision is incorporated at the ending of Elisabeth's second visionary book, see Elisabeth Schoenaugiensis, *Visionen*, 53–5.

[27] Clark, *Elisabeth of Schönau*, 40.

came about, the sources she was author of are the only recourse. And when the sources are slippery, the scholar must somehow position oneself vis-à-vis the design of these visionary texts, and partially make answers or even attributions.

Such answers often betray more about the scholar in question than about the author Elisabeth of Schönau. For Ferdinand W.E. Roth in the nineteenth century (the editor of Elisabeth's visions), distinguishing the author from the secretary would have been a matter of telling apart the sections 'most hysterical' within the visions in order to find the female voice.[28] The ecstatic, the dreamlike and the fanciful could not have been a male invention.[29] Indeed, the authorship of respectively masculine and feminine sections has been interpreted as a vital aspect by which the function and ordering principle of the visions can be gleaned. According to Marie-Geneviève Grossel, the apparition on the one hand and the ensuing gloss on the other can be regarded as the working spheres of respectively Elisabeth and Ekbert.[30] But this gendered division seems to answer only partially to an authorship problem that is much more complicated, and actually risks reinstalling stereotypical gender roles rather than refuting them. As Thalia Pandiri has emphasized in her contribution to a volume titled *Women Writing Latin*,[31] far more is at stake than a mere 'woman talking' vs. 'man talking' dynamic. Elisabeth's visions negotiate with and function within a Latin literary tradition where cross-gender interactions are easily imitated or artificially constructed. Pandiri, who set out to "decode" these subtexts,[32] brings up Abelard's *Historia Calamitatum* amongst other literary models to emphasize that autobiographical notions are to be approached with great caution in the twelfth century.[33] In this light, Heloise's letters responding to Abelard's *Historia* could just as easily be mentioned as a notorious example of a dialogue between man and woman which shows to have been meticulously constructed.[34]

### 5.1.3 The Corpus

The body of texts that appertain to the author Elisabeth of Schönau is complex. Internally, the authorial voices of Ekbert and Elisabeth seamlessly interweave, and form a patchwork of different genres (epistolography, hagiography, sermons, visionary trea-

[28] The prejudices toward medieval female writers and the denial of their authorship, amongst which also the questionable arguments of Roth mentioned here, are discussed more fully in chapter 1 on pp. 39ff.

[29] "Vor allem das innere Seelenleben, dass aus ihnen spricht, das Extase, die melancholischen Schmerzen, hysterischen Verzückungen, wie solches alles nur dem Weibe eigen ist [...] Dieses alles schützt die Visionen, deren einfache Angaben zu wahrheitsgetreu sind, um Mächwerk eines Mannes zu sein, vor allen Angriffen der Unächtheit," see Roth, *Visionen*, xcv.

[30] Grossel, "Voir des yeux du coeur," 367.

[31] Thalia A. Pandiri, "Autobiography or Autohagiography? Decoding the Subtext in the Visions of Elisabeth of Schönau," in *Women Writing Latin from Roman Antiquity to Early Modern Europe*, ed. Laurie J. Churchill, Phyllis R. Brown, and Jane E. Jeffrey, vol. 2, Women Writers of the World (New York: Routledge, 2002), 197–229.

[32] The term "decode" is found only in the title, and does not reappear elsewhere in the contribution.

[33] Pandiri, "Autobiography or Autohagiography?," 199.

[34] The letters of Heloise and Abelard are more extensively discussed from p. 235 onward.

tises ...). This lack of fixed structure is further reinforced on an external level, in the texts' transmission. Although one can distinguish recognizable, individual elements within Elisabeth's oeuvre, her texts were transmitted, edited and titled in various ways. We find pieces of Elisabeth's texts travelling independently, creating their own cycles, or adhering to collections that transcend the author level. There appears to have been no dominant *opera omnia* to stabilize the transmission. A comprehensive collection appeared only with Roth's edition in 1884.[35] Ekbert, who can be accounted for as the earliest of Elisabeth's compilers, is known to have drawn up compilations where some will have approached a type of *opera omnia*, but Ekbert equally disseminated Elisabeth's works in separate pieces. Clark noted that prior to Ekbert's arrival "there is no evidence of a formal collection of visionary texts published outside the walls of Schönau until at least 1159." This includes the visions which Elisabeth would have dictated to her fellow nuns some time around 1152.[36] The obvious result is that confusion might arise over which blocks of text correspond to what title, and who has most likely authored them. Add to this the false ascriptions made by the German Franciscan Livarius Oliger (1875–1950),[37] and disorder is complete.

One may be inclined to oppose this lack of systematicity within Elisabeth's corpus to Hildegard of Bingen's visionary treatises, whose integrity seems to have been compromised to a lesser degree. Not unlikely this is so because Hildegard herself had seen to prevent it by supervising her own *opera omnia*.[38] With Elisabeth, the tradition is rather one of bits and pieces of which we presume that they were partly written by Ekbert and partly by Elisabeth.[39] These were ultimately bundled to form a *Vita Sanctae Elisabeth* printed in the *AASS*, which would remain authoritative until the late nineteenth century.[40] This early edition of Elisabeth's works makes abundantly clear that scribal reproduction —and the inevitable variation the phenomenon brings along to the state of the text— continues well after the invention of print. Elisabeth's and Ekbert's writings in this *Vita* were arranged carefully to form yet another new composition, with a narrative that chronologically reconstructs Elisabeth's life. The title of *Vita* is therefore somewhat misleading in face of its contents. The collection of texts does not strictly contain a biography, nor is it replete with (auto)biographical accounts. Moreover, its division into ten chapters is an early modern invention,[41]

[35] Roth, *Visionen*.

[36] Clark, *Elisabeth of Schönau*, 45–6.

[37] Elisabeth Schoenaugiensis, "Revelationes B. Elisabeth. Disquisitio critica una cum textibus latino et catalaunensi," ed. Livarius Oliger, *Antonianum* 1 (1926): 52–82; Oliger's attribution of these texts was disputed by Köster, "Elisabeth von Schönau," 98–101.

[38] Or, at least, this is the working hypothesis. We return to this in the next chapter on p. 180, more specifically in n. 22.

[39] The most coherent structure within Elisabeth's oeuvre is given by her frequent reference to the catholic calendar, which seemed to have been a major inspiration and incitement for many of her revelations.

[40] *AASS* June III (16-19), Antwerp: Apud Viduam Henrici Thieullier, 1701, 607-643. Reprinted in: *PL* 195:119–94.

[41] The *AASS* print referred to in n. 40 had Lefèvre d'Étaples' edition as exemplar, see Jacques Lefèvre d'Étaples, ed., *Liber trium virorum & trium spiritualium virginum* (Paris: ex officina Henrici Stephani, 1513), 119–50; Köster indicated that this earliest edition was a major source for all later printed editions before Roth, in Köster, "Elisabeth von Schönau,"

and an artificial superimposition with a vague chronological principle underlying the compilation.

For centuries, these uncritical and fragmentary impressions constituted all the extant documentation for Elisabeth's texts. Contemporary scholarship mainly relies on later systematizations, notably Roth's edition[42] and Köster's catalogue of manuscripts,[43] which allow to orientate through the corpus. Unfortunately, Roth's edition is in spite of all its virtues still deficient. Although the edition is a big leap forward when compared to edition of the *AASS*, Köster (and later Clark) have signalized that it is based on an insufficient number of manuscripts, and manifests acute flaws.[44] However, despite Köster's critiques, the shortcomings of Roth's edition, and the collection of new manuscripts brought to light over the years, Clark has expressed confidence that "no major new discoveries of texts by Elisabeth will be made."[45] In awaiting a critical edition, the knowledge that at least all material is provided makes Roth's work stand as legitimate for our current purposes. Throughout the pages that are to follow, Roth's titles and structures will generally be respected unless indicated otherwise. This means —amongst other examples— that Elisabeth's popular visionary text of the *De Resurrectione* will remain integrated at the ending of the second visionary book, where Ekbert would have inserted it when assembling the collection, even though this text has a distinctive character from the remaining of the book and will often be found discussed separately. Kurt Köster's catalogue of "Benennungen Elisabetscher Schriften oder Schriftenteile," which was largely based on Roth, will be used to identify separate sections.[46]

By and large, the state of the art has agreed upon the catalogue tabularized in table 5.1. As will have been clear from the preceding paragraphs, this catalogue is based on convention amongst scholars. These visions are ascribed to Elisabeth (and these ascriptions will be respected), but Ekbert's influence in them is a matter of great dispute. So, for instance, the *Prologus* and *Narratio Eckberti* by which the first book of Elisabeth's visionary collection opens, have always been taken to be of Ekbert's hand. In table 5.1, however, they are nevertheless assigned to Elisabeth (as being implicit part of her visions) for conventional purposes.

245. It is unclear which manuscript(s) Lefèvre d'Étaples would have had access to and used for his edition.

[42] Roth, *Visionen*.

[43] Köster, "Elisabeth von Schönau."

[44] Roth mentioned nine complete and little more than twenty partial transmissions of Elisabeth's visions. This number has increased up to 145 known manuscripts (complete and partial). See Köster, "Das visionäre Werk Elisabeths von Schönau"; Clark has extended the list as well as corrected it, see n. 3 to her introduction in Clark, *Elisabeth of Schönau*, 149.

[45] Clark, *Elisabeth of Schönau*, 6.

[46] The abbreviations used by Köster also apply in table 5.1, see Köster, "Elisabeth von Schönau," 247.

| Author | Title (or *incipit*) | Abbrev. | Ed. (*R.* = Roth) |
|---|---|---|---|
| Elisabeth of Schönau | *Liber primus* | (*Vis.* I) | *R.* 1–39. |
| | *Liber secundus* | (*Vis.* II) | *R.* 40–55. |
| | *Liber tertius* | (*Vis.* III) | *R.* 56–87. |
| | *Liber viarum dei* | (*LVD*) | *R.* 88–122. |
| | *Liber revelationum* | (*Rev. Urs.*) | *R.* 123–38. |
| | *Elisabethae de sacro exercitu virginum Coloniensium* | | |
| | *Epistolae* (*Liber* VI) | (*Ep.*) | *R.* 139–53. |
| Ekbert of Schönau | *Magnificat anima mea* | (*MAM*) | *R.* 230–47. |
| | *Missus est angelus* | (*Miss.*) | *R.* 248–63 |
| | *De obitu* | (*Ob.*) | *R.* 263–78 |
| | *Soliloquium seu meditationes* | (*Med.*) | *R.* 278–303. |
| | *Epistolae Ekberti* | (*Ep. Egb.*) | *R.* 311–20. |
| | *Laudationes* | (*Laud.*) | *R.* 323–8. |
| | *Orationes* | (*Orat.*) | *R.* 338–41. |
| | *Ymnus de S. Gregorio* | (*Ymn.*) | *R.* 341–2. |
| | *Sermones XIII contra catharos* | (*Cath.*) | *PL* 195:11–102. |

Table 5.1: Corpus for Elisabeth and Ekbert of Schönau, sorted by appearance in Roth's edition (1884). The abbreviations for Elisabeth's texts are Köster's (1951). Abbreviations for Ekbert's texts are mine. Three shorter texts ascribed to Ekbert were not retained because they were useless for detecting stylistic patterns. They were mainly series of short laudatory salutations containing words such as *ave*, *o*, *gaude*, etc. The texts in question are the *Salutatio ad infantiam* (*R.* 320–22), the *Ad. S. Johannem B.* (*R.* 323) and the "Lobrede auf Maria" (*R.* 324).

## 5.2 Experimental Set-Up: Rolling *Impostors*

Now that we have gained a general impression of the corpus and its complex composition history, we consider an experimental set-up by which to tackle the problem of its double authorship. As table 5.1 suggests, we need to approach the case study of Elisabeth and Ekbert of Schönau as an authorship verification problem. There is no external reference corpus for Elisabeth, since none of the texts ascribed to her can be taken to represent her style. It could even be the case that the visions carry more of Ekbert's influence than of Elisabeth's. For Ekbert, then, we are in luck, since the secretary has been considerably prolific on his own. We dispose of a reference corpus by the latter consisting of $\pm$ 77,500 $w$ (the texts have already been tabularized in table 5.1). The question is how useful this reference corpus will be. A potential disadvantage is that Ekbert's corpus does not have any affinities to the visionary genre.

In short, the situation can be described as one in which a text of unknown authorship (Elisabeth's) is juxtaposed to texts of known authorship (Ekbert's), and we seek to verify whether or not these texts are in fact of same authorship or not. An experimental set-up was devised in which three demands rise.

**Demand 1** We need a method that is fine-grained. We need a method that can analyze

stylistic characteristics developing over short passages (Elisabeth's visions are ± 60,500 *w* long). For this we will use a so-called rolling method with sliding text windows.

**Demand 2** We need a quantification of the deviation from the visions' average patterns. For this we will use popularly used outlier detection methods, which have previously proven successful in intrinsic plagiarism detection. The underlying intuition is that the segments within the visions will display either 'normal' or 'aberrant' behaviour when compared to the visionary corpus as a whole. Segments, then, will be labelled as either compliant to expected behaviour, or as deviant from it.

**Demand 3** Ultimately, we will require a classification of the former. This means that measuring intrinsic stylistic variation as explained under condition 2 is all good and well, but such measurements by themselves mean very little if they are not explained. Also corpora of same-authorship are known to exhibit variability. For this condition to be respected, the *style change* gains an additional dimension against the benchmark corpus (A.I.1), which also includes Ekbert's texts. If some segment is consistently attributed to Ekbert in an *impostors* method (3.4.3, p. 117), the passage carries his influence. By contrast, if the segment is consistently labelled as 'None of the above' and does not meet the *impostors* threshold for Ekbert or any other author, we can presume a type of writing that is more 'Elisabethan.' A difficulty that arises in this scenario, however, is that such a classification failure is ambiguous. A segment that is 'unlike Ekbert,' does not necessarily make it univocally 'Elisabethan.'

### 5.2.1 Condition 1: Rolling Stylometry

In recent scholarship, the automatic detection of intrinsic stylistic interferences has drawn a great deal of attention. This is due to the fact that popular real-world applications of stylometry, for example the detection of plagiarized passages in scholarly theses, is almost always directed towards texts of dual or mixed authorship. These are scenarios in which there is no external reference corpus available, and where a given document stands on its own. For example, a student will have written the bulk of the text, but the source text(s) from which (s)he might have plagiarized can impossibly be known in advance. These dynamics are similar to those at play in Elisabeth of Schönau's visions, where we might assume that one stylistic signal is dominant, and that another will have occasionally interrupted the general patterns. The observation has arisen that slicing up the data into discrete chunks —or discrete sampling (A.3.3, pp. 318ff.)— carries much bias onto the decision-making. Since the text has been artificially sampled at will of the analyst's preset ranges, very little insight is gained as to how the text develops sequentially, and where to localize higher peaks of 'unexpected'

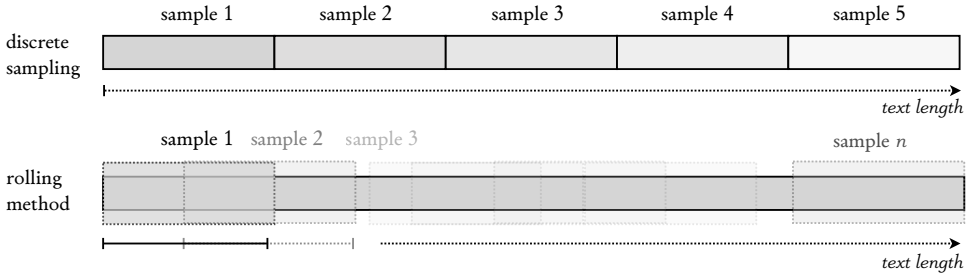



Figure 5.1: Intuition of discrete sampling (as had already been visualized on p. 92) vs. a rolling sampling method. The original text is sequentially sliced up in overlapping windows according to a step size. The philosophy of the method is that samples do not need to be extremely small in order to capture stylistic developments over a sequence of words.

patterns. Splitting up the text into smaller samples does not solve this problem, since smaller samples drastically decrease the accuracy of the attribution. For this reason, varieties of 'rolling' methods have gained leeway in the field. Their asset is that they sample the text in non-identical, partially overlapping windows instead of discrete chunks of text.[47] In such methods, the original sample size can be retained, and due to the linear overlap one is able to keep an eye on the neighbouring text passages. One can think of this as taking up a magnifying glass, scanning the text linearly, and registering how it changes from the very first to the very last word. In what is to follow, Elisabeth's visionary texts were sliced up into samples of 500 words, which is a very short sample size, but nevertheless worked surprisingly well in the sliding *impostors* method (explained more fully on p. 159).

### 5.2.2 Condition 2: Intrinsic Plagiarism Detection

Before setting out to benchmark Elisabeth's visions against a corpus of impostors, a quantification of how the individual windows behave in relation to one another might be of additional interest. This is an experiment in which we explore the 'intrinsic consistency' of Elisabeth's visionary corpus, and more importantly, in which we detect shorter sections that are outliers as opposed to the distributional properties of the larger document.[48] Outliers are considered to be 'loners' in a larger population:

[47] Maciej Eder, "Rolling Stylometry," *Digital Scholarship in the Humanities* 31, no. 3 (2016): 457–69.

[48] Intrinsic plagiarism detection is a rich field with a large literature of its own. A popular method for mapping the intrinsic style variation of a text has been Stamatatos' usage of character *n*-gram windows benchmarked against a vector representing the entire document. See Efstathios Stamatatos, "Intrinsic Plagiarism Detection Using Character *n*-gram Profiles," in *SEPLN 2009 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 09)*, ed. Benno Stein et al. (Universidad Politécnica de Valencia, September 2009), 38–46. A 'normal' profile is maintained, and each new window point is tested against this profile to decide whether it behaves as an outlier or not. In many aspects this is the idea of a centroid vector, an 'average' profile of the larger document, to which an individual window can either be similar (an *inlier*) or anomalous (an *outlier*). Although the idea remains generally valid, this particular method has in the meantime been criticized, and some suggestions have been offered to improve it, notably in Mike Kestemont, Kim Luyckx, and Walter Daelemans, "Intrinsic Plagiarism Detection Using Character Trigram Distance Scores under a Novel Document Representation," in *Notebook Papers of the 5th Evaluation Lab on Uncovering Plagiarism, Authorship and*

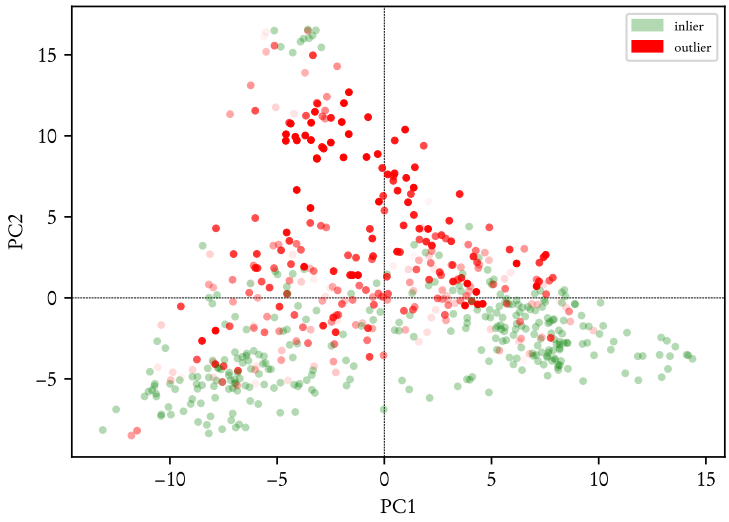



Figure 5.2: Intuition of outlier detection for Elisabeth of Schönau's visions. The intensity of red indicates the degree of 'anomaly.' As becomes clear from the plot, outliers are not necessarily located at the outer edges of a population. Some data points closely overlap, because the windows' contents overlap as well. Settings: *s–l* = 500 | *type* = most-frequent words | *n* = 1,000 | *vect.* = standard-scaled tfidf-weighted frequencies | *expl. var.* = 6.29%.

they are patterns that do not conform to expected stylistic behaviour. Under the presumption that the frequency of words respects a normal distribution, this means that outliers will have fewer neighbours, i.e. possess a lower density. Fig. 5.2 visualizes the window samples of Elisabeth of Schönau's visions as data points in a 2-dimensional space (principal components analysis or PCA). The data points are coloured either red or green so as to signify if they are out- or inliers. This PCA plot forms an intuition of the algorithm that we will be using to detect outliers, namely local outlier detection (LOF), which is an algorithm combining the intuition of *k* nearest neighbours and outlier detection.[49] A common misunderstanding —probably resulting from the term's name— is that an outlier lies 'outside' of the data's distribution. Although this actually can be true, it is not a general rule. The power of the LOF algorithm is that such 'global' outlier characteristics are traded in for 'local' properties, which is the virtue of implementing a *k* nearest neighbours approach. To fully understand the forces at play in fig. 5.2 a portion of the appendix at the back of this thesis is dedicated to the LOF algorithm (pp. 361ff.).

*Social Software Misuse (PAN)* (Amsterdam, 2011), Comparing the profile of a longer text (the whole document) against that of a shorter text (the window) is theoretically counter-intuitive, since the lexical range of the pair will be unequal. Moreover, Stamatatos' approach comes with the implication that the suspicious document represents a text genuinely written by a single author, with only a handful of small intrusions. A more flexible method would be to calculate the distances between respective window vectors in a type of nearest neighbour approach (see A.3.5, p. 335). Consequently, the window vectors that behave anomalously as opposed to the entire corpus can be traced by using outlier detection methods.

[49] Markus M. Breunig et al., "LOF: Identifying Density-Based Local Outliers," *ACM SIGMOD Record* 29, no. 2 (2000): 93–104.

### 5.2.3 Condition 3: *Impostors* Threshold

Condition 3 entails that each of the sliding windows receives a label of either 'Ekbert,' or else 'not Ekbert,' i.e. the authorship verification problem. For this complex question in computational authorship attribution, the *impostors* method and its variants is considered the finest state-of-the-art solution.[50] Important to note at this point is that we are interested also in quantifying the degree of stylistic influence on the text sample, perhaps even more than in merely its 'actual' authorship. If any influence by Ekbert is betrayed at all, we would like to see this reflected in a quantification of degree (e.g. a percentage). 500-word samples written by Ekbert were compared to himself and a group of 24 impostor authors from the benchmark corpus (see table A.I.1 on pp. 304ff.), in order to observe how well these samples adhere to his authorship in the framework of a larger group of authors. It turned out that in the traditional *impostors* method set-up achieving correct classifications was possible, whereas training the $\sigma^*$ threshold was very hard. Whereas I could tune the algorithm's parameters well enough to make many correct attributions, it was near to impossible to make it confident of these attributions by outputting high $\sigma^*$ scores.[51] Therefore, the 'None of the above'-label, which is typical to authorship verification methods, failed to optimize, as also <*diff-author*> pairs kept transgressing it.

It should be noted that the *impostors* method as proposed by Koppel and Winter is typically a very strict method, which only attributes a text sample when it is very confident of attribution. I therefore propose a more relaxed variation upon the *impostors* method which has trained on only one class, i.e. 500-word samples from Ekbert's corpus —as tabularized in table 5.1— and was played out against each of the candidate impostors from the benchmark corpus.[52] For the time being, this mild and relaxed variation on the *impostors* method succeeded to classify Ekbert's works to himself instead of to any of the potential impostors at a threshold of 0.20 and an accuracy of 91.95% (see fig. 5.3), which was to some extent surprising in light of the samples' short length. This means that 500-word window samples can in fact be reliable enough to detect sections within Elisabeth's work that are very 'Ekbertian' as opposed to 24 impostor authors. Within the current framework, where we know on a historical basis that Ekbert was involved in collaborating on Elisabeth's visions, this should be sufficient.

[50] The *impostors* method was explained in full in 3.4.3 (p. 117).

[51] The technical details of the *impostors* method are more elaborately explained in chapter 3. On p. 119 I have noted that —for some historical texts and for some authors in particular— finding a steady $\sigma^*$ threshold can be difficult or even impossible.

[52] The corpus was segmented into 500-word samples, each of which were vectorized (tfidf) and scaled (standard) by a random feature subset consisting of 50% of the corpus vocabulary.

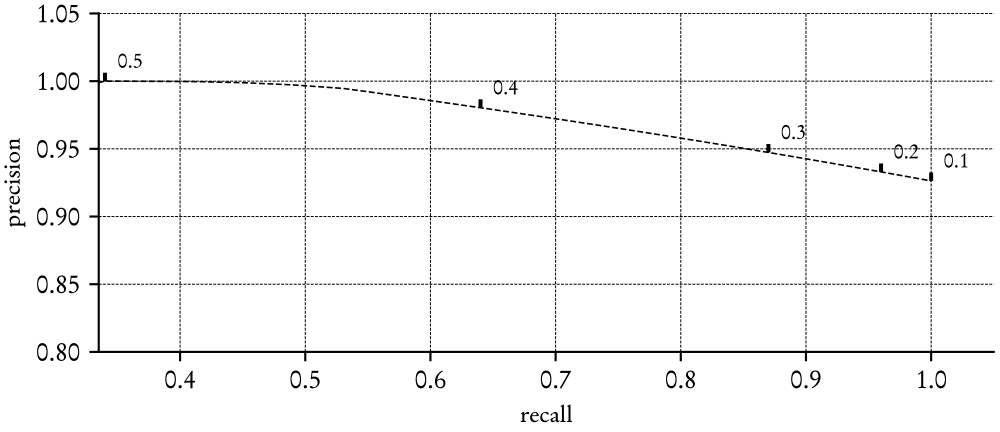



Figure 5.3: PR-curve for the *impostors* method with different $\sigma^*$ thresholds in a preliminary training round for 500-word samples of Ekbert of Schönau (tfidf-weighted vectors with standard-scaling).

### 5.2.4 Sliding *Impostors* Method

The result is a model in which all three preceding conditions are represented. Firstly, it requires the 'rolling windows' preprocessing aspect in order to make more fine-grained analyses of style developing as the text unfolds. Consequently, each of these windows' stylistic profile *within* Elisabeth's entire visionary corpus needs to be determined. The aim is to uncover whether or not a particular window betrays any anomalous behaviour in relation to all neighbouring passages (i.e. outlier detection). Thirdly, then, each of these windows' stylistic profile *without* Elisabeth's visionary corpus is determined. In this final condition, the *impostors* algorithm is applied to test the text window against 24 impostors and one author of special interest: Ekbert. We have shown that distinguishing Ekbert from this set of impostors achieves considerably high accuracy (91.95%) despite the short sample length (fig. 5.3, p. 160). Therefore, we can be confident that sections within Elisabeth's visions assigned to Ekbert and transgressing the $\sigma^*$ threshold of 0.20 can indeed be attributed to Ekbert by a high degree of confidence. The power of this experimental set-up, which will henceforth be called the 'sliding *impostors* method,' is that all of these significant aspects are combined into one visualization. The intrinsic plagiarism detection (outlier detection) in the backdrop of the *impostors* method allows that these two algorithmic sources of information shed mutual light on each of both.

## 5.3 Results

### 5.3.1 A Bird's Eye View of Elisabeth's Visionary Diaries

Fig. 5.4 (p. 162) paints a vivid picture of cross-gender collaborative authorship in the twelfth century. In the figure, the 500-word window samples unfold by steps of

100 words towards the ending of the visionary corpus, and for each step the LOF and *impostors* algorithm make their predictions. This bird's eye view with its various colour palette might be somewhat intimidating at first glance. Therefore, we will first discuss some of the larger, rough trends (a true 'distant reading'), before delving into some of the plot's subtrends in more detail. I will refrain here from elaborating on all of the technical principles behind the figure, which are described within the figure's caption on p. 162.

It should be emphasized that outlier detection (intrinsic plagiarism) in reds, and the *impostors* method with respective attributions to Ekbert (oranges) or others (greens), are very different measurements. Nevertheless, it is striking to see that —on a distant level— there is a rough match to be discerned in both algorithms' results. Sections of which the *impostors* method is very confident that they belong to Ekbert (dark oranges), are often classified as 'outliers' within the visionary corpus (red peaks). Additional faith in the current method is gained by observing that the experiments confirm some of the relevant state-of-the-art assumptions when it comes to Ekbert's stylistic influence. The *Prologus* and the *Narratio Ekberti*, for instance, standing at the very opening of the corpus, are attributed to Ekbert (the two texts are very short and together compose the very first 500 words). In addition to this stylistic affiliation with Ekbert, this short introductory section to Elisabeth's visionary corpus constitutes an outlier.

Secondly, what jumps in the eye in this preliminary distant reading, is that Ekbert's influence over Elisabeth's literary corpus is generally very large, be it in some sections statistically more convincing than in others. It was to be expected that some stylistic affinity with Ekbert would have shown itself. The extent by which, however, is quite surprising. Ekbert's Latin surfaces dominantly in all of Elisabeth's visions, even though the earlier visionary material was first recorded between 1152–4, some time before Ekbert's arrival in 1155. This need not be surprising, as Clark's study of the manuscripts led her to the conclusion that the final collection as we have it was mostly a compilation by Ekbert, including the latter's additions and revisions to Elisabeth's older works.[53] What becomes clear is the complexity of this collaborative writing process, and the extent by which the voices of Ekbert, Elisabeth and (possibly) other collaborators are entwined. The figure allows to trace how a sheer number of paragraphs (some 100 words) can throw a sample's profile out of balance completely.

Thirdly, although there is no obvious chronological trend emphasized within the figure (the works are nevertheless presented in their assumed chronological order), one can discern that Ekbert's style is far less dominant in Elisabeth's first visionary book. Greens ('not by Ekbert') are better represented. From *Vis.* II onward, which was the first book on which Ekbert collaborated, the orange and red tints gain in pres-

[53] Clark, *Elisabeth of Schönau*, 46ff. Most of Ekberts revisions to the texts took place after Elisabeth's death. See 131.

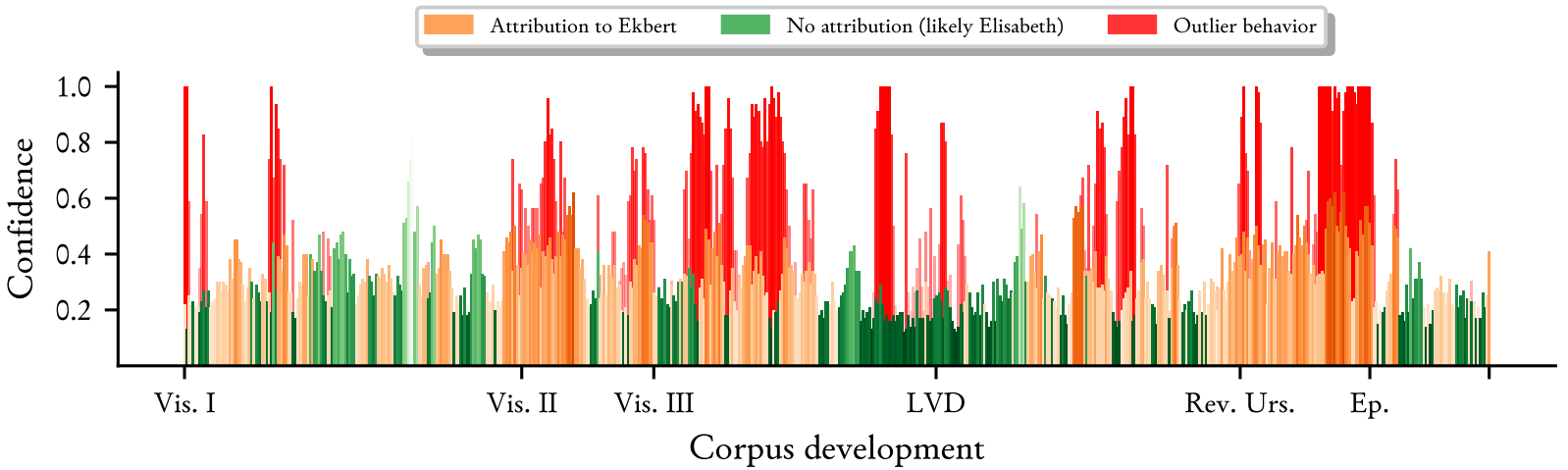
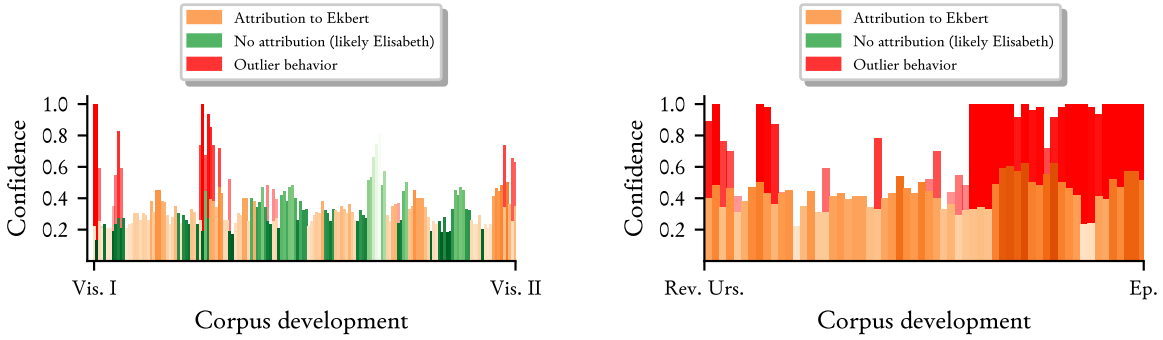



Figure 5.4: Results of the sliding *impostors* method for Elisabeth of Schönau's visionary corpus. The x-axis shows the progression of the sliding windows throughout Elisabeth's corpus by taking a step of 100 words at a time. In total, attributions were made for 601 partially overlapping text samples consisting of 500 words. The x-axis labels indicate where each of Elisabeth's books begins (abbreviations in table 5.1). The y-axis (the height of the bars) indicates the min-max-normalized confidence scores (between a range of 0 and 1) for both the LOF algorithm as the *impostors* method. Reds indicate that the sample is considered an outlier. The colours' intensity and the confidences (y-axis) betray the degree by which. Oranges indicate that a text window of 500 words is attributed to Ekbert, and the *impostors* confidence threshold ($\sigma^*$) of 0.20 is surpassed. The colour intensity increases as $\sigma^*$ contains a higher value. Note that for some sections the $\sigma^*$ threshold of 0.5 is surpassed, which is the point at which precision during the training phase had reached 100%. Greens indicate an attribution to any of the impostors (green), which we know must be coincidental and therefore receives the label of 'no attribution (likely Elisabeth).' A contrasting logic than for Ekbert was adopted for colour intensity. If the confidence $\sigma^*$ for one writer amongst all impostor candidates is very high, this is more likely to stand further apart from Elisabeth's writing. This explains why darker greens conform to a lower $\sigma^*$ (many of which remain under threshold 0.2). Greens, in general, do not achieve a very high $\sigma^*$, since they generally indicate that the sample lacks a convincing impostor profile to whom the sample can be attributed. Lighter greens are very confident attributions to authors other than Ekbert, which are interesting in their own right, but can in the current scenario only be seen as coincidental output which is not easy to explain.




(a) Detail of Elisabeth of Schönau's first visionary book (*Vis.* I), written c. 1152.

(b) Detail of Elisabeth of Schönau's revelations about Saint Ursula (*Rev. Urs.*), written c. 1156/7.

Figure 5.5: Details of fig. 5.4, exemplifying the drastic difference in style and the obvious intrusions by Ekbert in Elisabeth's first visionary book and her last. Note that both visions differ in length (±15,750 $w$ vs. ±5600 $w$), hence the wider strokes on the right-hand figure.

ence.[54] The discrepancy between Elisabeth's older and younger writings and Ekbert's increasing number of intrusions becomes drastically clear when comparing the profile of Elisabeth's *Vis.* I (left) with her *Rev. Urs.* (right) in detail (fig. 5.5). Elisabeth's debut, which must have been first drawn up c. 1152, has a completely different authorial profile than her last book c. 1156/7.[55] This last work, containing Elisabeth's revelations about Saint Ursula and her eleven thousand virgin companions, suggests little trace of 'another' style than Ekbert's, nor of some synergetic effort. From a statistical perspective, there is little reason to doubt that the *Rev. Urs.* is in fact drawn up in Ekbert's Latin. This raises a question or two about Elisabeth's intensifying collaboration with her brother and the dynamics between the two. After all, Elisabeth's last book on Saint Ursula and her eleven thousand virgins is not just another vision book within her oeuvre. As we will discuss below, it is a book that is politically charged, and shows direct involvement with ecclesiastical controversies in the world at large, i.e. the world outside Schönau governed by men.

### 5.3.2 Identifying Elisabeth?

The sliding *impostors* method in fig. 5.4 gives an indication of which samples are most 'Ekbertian' (in orange). The green samples, on the other hand, only have in common that they are unlike Ekbert and —if they have a low confidence— that they are unlike any of the impostors. It would be interesting to ascertain whether or not there is method in this aberrant stylistic behaviour, and if the green samples that remain unrecognized show a similar stylistic profile with regards to each other. After

[54] Köster, "Das visionäre Werk Elisabeths von Schönau," 83.

[55] Köster derives the dating from Schönau's reception of the Ursuline relics coming from Cologne, which indeed must have been around the year 1156/7, see ibid., 81–2.

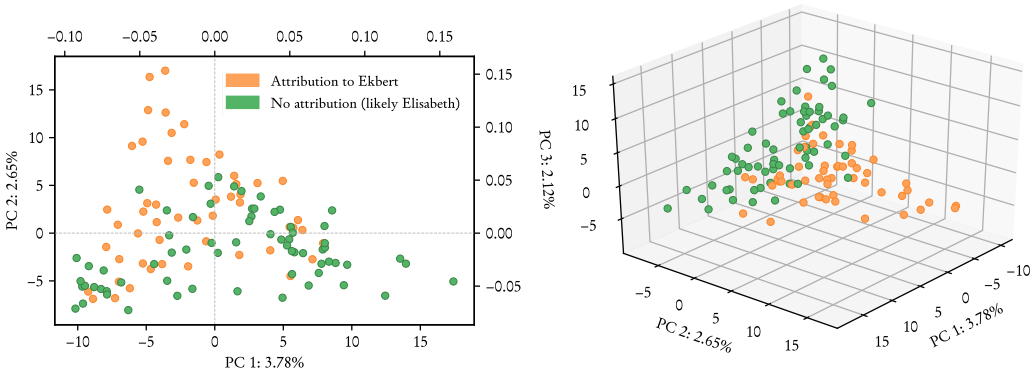



Figure 5.6: PCA plots of Elisabeth's visions, visualizing discrete samples attributed to Ekbert (orange), and those attributed to the distractor authors in the *impostors* method (green) (fig. 5.4). The first two principal components suggest a subtle clustering of samples that belong to Ekbert as opposed to the ones attributed to impostors (which therefore may likely be more 'Elisabethan'). Especially in the right-hand, three-dimensional PCA plot, this clustering becomes more convincing when a third PC is introduced.

all, if all of the green samples would show resemblance to each other, the likelihood of their singular authorship increases, which in its own turn strengthens the hypothesis that some sections in the visions are the work of Elisabeth and Elisabeth alone.

## PCA

One way to explore this question is by assigning labels and colours to the samples that were attributed to Ekbert and to those that were unidentified, before consequently visualizing them in a PCA plot (fig. 5.6). The PCA plots (both two- as three-dimensional) make clear, however, that distinguishing between the samples is not at all easy. All of the samples cluster together quite intensively. However, there appears to be a very vague plain distinguishing between both groups of samples. In the left-hand PCA plot, Ekbert's samples are somewhat better represented in the upper half of the figure (along the second component) than the green samples. In the three-dimensional plot on the right-hand side of fig. 5.6, an even better plain presents itself along which the samples appear separable. The green samples group together at the left of Ekbert's orange samples. In the end, however, despite their suggestions of a subtle distinction between both groups, the PCA plots do not suffice to give any conclusive results on different authorship, which may well be the reality of this text.

## A Second *Impostors* Method

A more powerful method is needed, and we need not look far. Determining if the green samples are more similar to each other than to other samples boils down to attempting to determine if two documents are written by the same author, which is —again— the authorship verification problem. Therefore, we simply repeat here

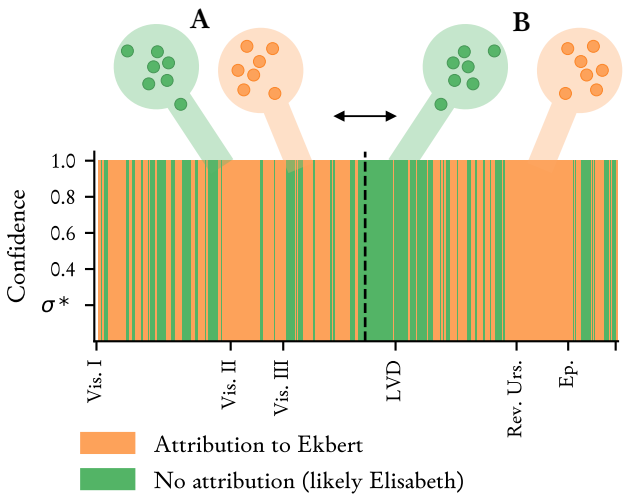



Figure 5.7: Batches A and B of Elisabeth's visions for text-internal comparison. The division held into account that an equal number of green samples was maintained (104 per batch) during the rolling *impostors* method. Samples from within the same batch were never compared.

the rolling *impostors* method, whilst taking into account one major game changer. Similarly as in fig. 5.4, we roll over Elisabeth's visions by taking samples of 500 words and a step size of 100 words. Again each of the samples is compared to Ekbert's texts (table 5.1) and the impostors from the benchmark corpus. However, aside from this habitual pool of documents, the sample is also compared to discrete samples taken from the visionary works themselves. This latter set of additional benchmark documents comprises two distinct types: samples that were attributed to Ekbert (orange) in the first *impostors* method, and samples that had been labelled 'unknown' (green). The goal is to test if the green samples, when forced to choose between 'green' and 'orange' companions, feel more inclined to adhere to the former than to the latter or vice versa.

The greatest problem with this set-up was that lexical overlap or content-dependent items can heavily bias the results. We therefore decided to split up the visionary material in two batches (A and B in fig. 5.7). This prevented that green samples from group A were compared to neighbouring green samples, and were always compared to the other half of green samples preserved in batch B. Both groups (which consisted of 104 samples each) in turns served as the respective test or impostors set, thereby simulating a cross-text approach.

The results of this second *impostors* method proved interesting, as the majority of green samples were not only unlike Ekbert, but also often like each other. As one can see visualized in fig. 5.8, approximately 60% of the green samples preferred 'green companions' over orange ones. The remaining 40% of the green samples were won over to the side of impostor documents (8.17%), or to the stylistic profile of orange samples (33.17%). However, even though 60% is a majority, it is hardly an overwhelming one, wherefore simplifications are to be shunned. That some 40% of

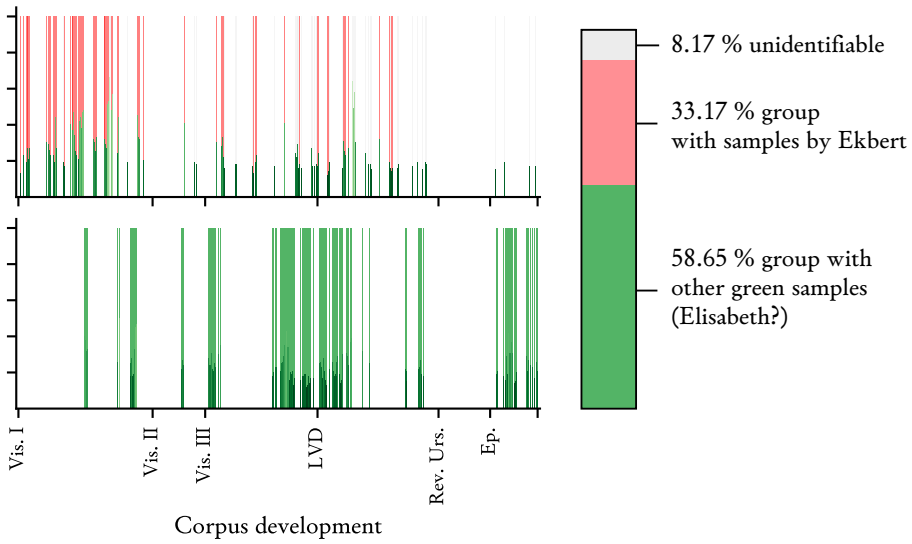



Figure 5.8: Visualization of the second *impostors* method, testing if the samples that are unidentifiable (green in fig. 5.4) and labelled to be unlike Ekbert have a stylistic similarity with each other. The second *impostors* method confirms that the majority of green samples indeed seeks out each other's company, even though they come from very different sections in Elisabeth's visions (for the precautions taken to prevent bias by neighbouring samples, see fig. 5.7). On the other hand, that some 40% of the green samples better resemble the samples attributed to Ekbert, or are taken in by impostor documents, raises cautions against an impulsive equation of the green samples with Elisabeth's hand.

the green samples better resemble the samples attributed to Ekbert, or are taken in by impostor documents, raises cautions against an all too impulsive equation of the green samples with Elisabeth's hand. There appears to be the suggestion that quite a few samples float in between the two, or else testify of a unique character of their own which cannot be accounted for easily.

### 5.3.3 The Cathars and the Ending of Book III

In the remaining paragraphs of this chapter fig. 5.4 is scrutinized in some more detail. The first suspect passage to which we will turn our focus is the ending of *Vis.* III. The section, which specifically begins from §22 onward,[56] is an elaborate condemnation of the Cathars.[57] Together with Hildegard of Bingen, Everwin of Steinfeld († 1152) and Bernard of Clairvaux,[58] Elisabeth and Ekbert have often been read as early sources on

[56] Elisabeth Schoenaugiensis, *Visionen*, 75.

[57] Elisabeth also condemns the Cathars elsewhere for their celibacy in her *Liber viarum dei*, more specifically in the *Sermo tertius de via coniugatorum*, see Elisabeth Schoenaugiensis, *Visionen*, §13, 104: "Domine mi, quid dicis de illis, quos katharos vocant, qui vitam coniugatorum omnino reprobare dicuntur?"; This translates to "My lord, what do you have to say about those who are called Cathars, who are said to completely condemn the life of married people?" Translation taken from Elisabeth Schoenaugiensis, *Complete Works*, 183.

[58] Beverly Mayne Kienzle, "Defending the Lord's Vineyard: Hildegard of Bingen's Preaching against the Cathars," in *Medieval Monastic Preaching*, ed. Carolyn Muessig, Brill's Studies in Intellectual History 90 (Leiden: Brill, 1998), 163;

the 'Catharist' doctrine coming to the rise in the Rhineland around the mid-twelfth century.[59] In neighbouring cities and particularly in Cologne and Bonn —where both Ekbert and Elisabeth had close connections—[60] the sectarian reform movement had been gaining traction. Ekbert had been concerned by what he termed *cathari* some time before he joined his sister in Schönau (1155), and it has been argued that "his interest probably fueled Elisabeth's."[61] However, the opposite has also been claimed.[62] In 1163/4, Ekbert would write a book of thirteen sermons against these dissident groups increasing in numbers in the Rhineland,[63] shortly after they had been publicly condemned and burned at the stake in Cologne in the summer of that same year. For these executions, Ekbert had been summoned as a kind of 'proto-inquisitor.'[64]

In Elisabeth's invective, the Catharists' rejection of Christ's incarnation takes up a central position. The passage sets out with a letter to Hildegard in which she describes the dualist heretics as serpents coming into the Church of God.[65] This leads into an excursion on human pride and sinfulness, not coincidentally through the voice of Christ himself in the first person. Consequently, Elisabeth accuses the Church of failing to take action against the Cathars' doctrines. From §29 onward, Elisabeth recounts

and Beverly Mayne Kienzle, "Crisis and Charismatic Authority in Hildegard of Bingen's Preaching against the Cathars," in *Charisma and Religious Authority: Jewish, Christian, and Muslim Preaching, 1200-1500*, ed. Katherine L. Jansen and Miri Rubin, Europa Sacra 4 (Turnhout: Brepols, 2010), 73–91.

[59] What exactly *is* Catharism, and whether or not it constitutes a movement or instead a discursive invention of the Church to target at once a number of dissenting dualist teachings and scriptural interpretations which it argued to counter its authority, has been subject to much debate. The traditional description of Catharism was that it had been a heterodox and heretical movement with roots in the Byzantine East and associations with Bogomilism in the Balkan which had flown over from the East. Three of its most controversial tenets would have been dualism (the existence of two Gods, one good and the other evil), the rejection of the catholic sacraments, and the denial of Jesus Christ's incarnation. This uniformity of a categorizable and identifiable 'Cathar' movement, and its indebtedness to Eastern cults has increasingly come to stand under question by scholars, and has been dismantled for being a polemic construction by later commentators and inquisitors. Brunn has shown that Ekbert may in fact have played an important role in 'coining' Catharism as a term to cover for the kinds of beliefs he suspected them to foster. See Uwe Brunn, *Des contestataires aux «cathares». Discours de réforme et propagande antihérétique dans les pays du Rhin et de la Meuse avant l'Inquisition*, Collection des études augustiniennes. Série Moyen-Âge et Temps modernes 41 (Paris: Institut d'études augustiniennes, 2006), see the introduction 13–29; and see 315–42 for Ekbert's importance in making the association between what he argued to be contemporary Catharism with the Manichaean *cathari* in late antiquity. Also see Moore's dismantling of Catharism in Robert I. Moore, *The War on Heresy. Faith and Power in Medieval Europe* (London: Profile Books, 2012).

[60] We will come back to this point from p. 170 onward.

[61] Clark, *Elisabeth of Schönau*, 22.

[62] "That Elisabeth had a strong personal influence upon Ekbert is clear enough therefore from the visionary collection. It would be surprising if his close connection with Elisabeth had not affected him, and his biblically erudite *Sermones contra Catharos*, written around 1164, suggest one way in particular in which it may have done so,"see Coakley, *Women, Men and Spiritual Power*, 33.

[63] The *Sermones contra catharos*. See *PL* 195:11–97.

[64] The death sentences took place on 5 August 1163, and Ekbert had been summoned by the archbishop of Cologne to dispute with the Cathars. See Robert I. Moore, *The Origins of European Dissent*, Medieval Academy Reprints for Teaching 30 (Toronto: University of Toronto Press, 1994 (1977)), 181; Thomas H. Bestul, *Texts of the Passion: Latin Devotional Literature and Medieval Society* (Philadelphia: University of Pennsylvania Press, 1996), 80–2; Brunn, *Des contestataires aux «cathares»*, 197–239.

[65] Around 1163, when anxieties concerning heresy arose again, Hildegard would set down her quill against the Cathars as well. She wrote two letters, one treatise, and one *expositio* or homily. It is thought that Elisabeth inspired her to do so. For a full study on this topic, see Kienzle, "Defending the Lord's Vineyard."

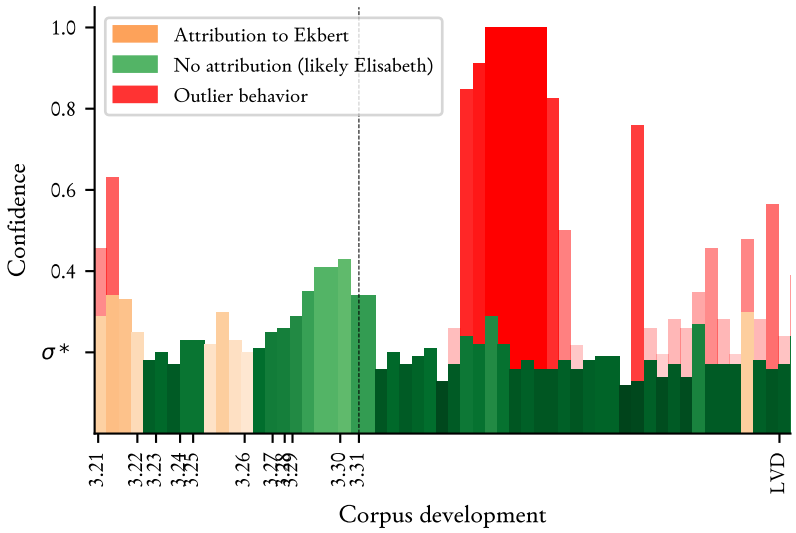



Figure 5.9: Detail of fig. 5.4, zoomed in on the final chapters of *Vis.* III, containing Elisabeth's criticism on the Cathars. The dotted, vertical line indicates the moment in the text at which Ekbert's commentary is announced. Strangely enough, Ekbert's exposition to his sister's vision has no stylistic bearing on the text. The red peaks, however, indicate outlier behaviour, or a stylistic presence that is considered aberrant throughout all of Elisabeth's diaries.

how she was seized by a vision some time during Lent when Ekbert happened to be absent.[66] After having described the vision, she entreats "her most beloved brother" to "examine the divine scriptures and try to discover a suitable interpretation of this vision," at which point Ekbert literally enters the text in §31.[67] After a topos of modesty in which Ekbert declares himself "insignificant" (*pusillus*) and "of little capacity" (*exigui sensus*) to undertake such a task, he accepts his sister's request and brings out a lengthy analysis of her vision, for which he consults a series of scriptural authorities. The ending of *Vis.* III comes full circle when Elisabeth's vision is interpreted as portending the fundamental importance of Christ's incarnation and sacrifice for the catholic Church, the exact aspect which the Cathars dare call into question.

The passages just described practically contain all the ingredients to suspect Ekbert's intrusion. Indeed, as mentioned above, there have been those who have suspected that Elisabeth wrote the passages on the Cathars at Ekbert's instigation.[68] Moreover, Ekbert is here explicitly designated author of the text from §31 onward. And yet, despite of what may appear to be obvious cues, there is no trace of Ekbert's stylistic signal in these passages. On the contrary. It is exceptionally absent when glancing upon the larger trends in the whole picture (fig. 5.4). A detail of this figure, given

[66] Elisabeth Schoenaugiensis, *Visionen*, 78.

[67] "Nunc igitur amantissime frater hunc tibi laborem, queso, assume, ut divinas scripturas scruteris, et congruam ex eis interpretationem visionis huius coneris invenire, tibi enim fortassis a domino reservata est," in Elisabeth Schoenaugiensis, *Visionen*, §30, 79. Translations taken from Elisabeth Schoenaugiensis, *Complete Works*, 149.

[68] Clark, *Elisabeth of Schönau*, 22.

in fig. 5.9, demonstrates that the ending of *Vis.* III is largely an unrecognizable zone within Elisabeth's entire oeuvre. It is coloured green, has a low confidence score, and shows little trace of Ekbert aside from a few occasional and unconfident attributions in the beginning. The red peaks in the middle of the passage moreover testify of these passages' unusual stylistic patterns, which is the only time throughout all of Elisabeth's visions that green-coloured 'unrecognized' attributions are signalized by the LOF algorithm as outliers. In short, these fragments are not only unlike Ekbert, they are unlike anything else the algorithm has observed throughout the rest of the corpus. Strangely enough this somewhat goes against the intuitions of scholarly consensus. Coakley, for instance, in following the logical flow of the narrative in which the speaking subject shifts from Elisabeth to Ekbert, argued that "Ekbert's erudition is most conspicuous at the end of the third *Liber visionum*."[69] In light of the fact that —at least, on a stylistic level—, such a shift does not seem to take place, we need to reassess again the literary techniques at play in this complex corpus of texts. Could these instead be the most 'Elisabethan' passages throughout the entire visionary corpus? Does it confirm Clark's argument that Elisabeth's integration of the Cathars in her own texts was different "in style and message"?[70] And do these results provide evidence that the interplay of roles as suggested by the text itself should not simply be taken for granted?

One aspect of Ekbert's exposition in particular considerably hardens a monocausal answer to these questions. As also Coakley did not fail to notice, "scriptural allusions appear by the dozens" in how Elisabeth's vision is interpreted.[71] Whoever authored the *expositio* had heavily researched authoritative sources, whose presence in their own turn might have thrown the stylistic signal out of balance as much as it does in fig. 5.9. That being said, I have reservations if intertextuality is a sufficient explanation. Firstly, even though the number of citations is high, it is not exceptionally high. There have been other samples by Ekbert replete with intertextual citations which thusfar presented no hindrance, nor has intertextuality presented a major concern in other texts analyzed in this thesis. Moreover, the majority of the *expositio*'s content still consists of the stylistic 'stuffing' to connect the citations, i.e. written material certainly drawn up by a Schönau author —Ekbert? Elisabeth? Some other scribe?— to interlace the allusions. It is also not the case that a label of 'No attribution' (in green) shows a correlation with the presence of intertextual references per se, as the preceding samples (those before §31 in fig. 5.9) barely have them, and are nevertheless coloured green and considered to be unlike Ekbert's style. But the most conclusive evidence that these samples are not thrown out of balance by intertextual allusion, is the experiment as carried out in fig. 5.8, where we illustrated that a small 60% of the green samples in fact show similar stylistic behaviour. As one will be able to verify

[69] Coakley, *Women, Men and Spiritual Power*, 29.

[70] Clark, "Repression or Collaboration?," 161.

[71] Coakley, *Women, Men and Spiritual Power*, 29.

there, this holds true for the ending of *Vis.* III.

The observation that these passages show so few traces of stylistic change in favour of Ekbert even when he is explicitly figured as respondent to Elisabeth's prompts should heed warning. This indicates that not only Elisabeth and Ekbert's composition, but also the representation of their composition process, is a construct, which must be approached with the utmost care. The way in which Ekbert's authority and authorship is conjured up by Elisabeth has a function and purpose on a textual level rather than a historical level. The discrepancy that shows on a stylistic level (and we will discuss another one below) prompts new reflection on how Ekbert and Elisabeth's role-playing as re-enacted in the text cannot simply be taken to coincide with the roles taken up in the material and actual setting of their collaboration. The passage's conspicuous divergence away from Ekbert's usual style, and its correspondence to other unrecognized green samples within the visionary works (fig. 5.8), imply that Ekbert's response to Elisabeth's vision was integrally the work of another author, by all likelihood Elisabeth herself. The question is now how we are to imagine the composition of such a passage. Perhaps Elisabeth and Ekbert sat down to ponder upon the meaning of the vision, and drew up the *expositio* in the kind of conversational mode in which it has often been presumed that their composition processes went about. In this more informal and creative setting, Elisabeth might have indicated to her brother which scriptural passages she herself had been thinking of, and she may even have gone as far as to explain in her own terms why she believed the passages were apt for explaining the vision. Moreover, that the passage on the Cathars shows little trace of Ekbert's style, allows due reconsideration: was Ekbert or Elisabeth most fervent in the resistance against the heretic movement?

### 5.3.4 The Colognian Connection and the *Book of Revelations*

Before discussing the next suspect passage in Elisabeth's visionary works, a concise introduction to the Ursuline legend and its revival in the twelfth century is called for.[72] Saint Ursula was a legendary fourth-century king's daughter from Brittany († 383). The tale goes that upon her betrothal to the heathen king of England by her father, Ursula would only accept the marriage under condition that she was first endowed a companionship of eleven thousand virgins,[73] with whom she made a pilgrimage to the city of Rome. This would safeguard the prosperity of both spouses and their

[72] A classic in exploring the origins of the legend of saint Ursula is Wilhelm Levison, *Das Werden der Ursula-Legende*, in *Bonner Jahrbücher* 132 (1972), 1–164. For a more recent volume of various studies on the legend and its cult in the Middle Ages, see Jane Cartwright, ed., *The Cult of St Ursula and the 11,000 Virgins* (Cardiff: University of Wales Press, 2016).

[73] That Ursula's company consisted of such a large number as eleven thousand, was likely to have been a misreading of the Roman numeral $\overline{\text{XI}}$. The stroke on top of the characters could either indicate that the characters were to be interpreted as numbers (hence, 'eleven'), or else that they were to be multiplied by thousand ('eleven thousand'). Confusion between both conventions caused the myth to change radically. An earlier suggestion had been that XI M was interpreted by a scribe to mean 'undecim milia' instead of 'undecim martyres.' See Levison, *Das Werden der Ursula-Legende*, 37–40.

marriage, and moreover advance the conversion of her future husband to Christian faith. However, as Ursula returned from her pilgrimage to Rome with her holy band, the companionship kept post for some time in Cologne. Their temporary stay there was disturbed when the Huns, under command of Attila, besieged the city. When Ursula refused Attila's advances to marry him, the angered leader of the Huns pierced her with an arrow, upon which she became a martyr of Christendom.

In 1106 and the decades to follow, the Ursuline legend would go through a revival, and become one of the most popular saints' cults of the medieval period. An excavation in the northern perimeter of Cologne (of which Schönau lies some 130 kms southwest) had come with the discovery of ancient Roman tombs containing bones of Ursula's companions.[74] The find incited the rise of an Ursuline cult which continued to flourish up until the fifteenth century. It also positively influenced a profitable bone business in Cologne, as the city attracted pilgrims and as the virgins' holy remains were trafficked across the continent. The initial exuberance over the thrilling find was, however, disrupted by findings of bodies of children and men, and of unattested names inscribed on stones within the graves. This sudden appearance of male characters in the legend showed inconvenient inconsistencies with the versions of the Ursuline story at large.[75] It is well known that around the half of the twelfth century increased skepticism arose concerning forgeries and other falsities. The acquisition (and vending) of relics was a serious and economic matter, and so was the crime and fraud associated with them. The Ursuline relics were no exception, and some fifty years after their initial discovery could not escape renewed suspicion.

Gerlach, the abbot of Deutz, appealed to Elisabeth's visionary gifts in the hope that some identification of the remains could be established. Elisabeth emphasized in the fourth chapter of the *Book* that the Colognian abbot fostered suspicions "that the discoverers of the holy bodies might have craftily had those titles inscribed for profit."[76] In his resolution to set facts straight, Gerlach transferred two of the unearthed bodies to Schönau, one a maiden by name of Verena, the other one of the virgins' cousins by name of Caesarius (a man). It does not seem as if Gerlach's request was initially met by great enthusiasm, as Elisabeth stated that she was "very resistant," until "certain men of good repute pressed me with their demand to investigate these things at length and they [did] not allow me to be silent."[77] Indeed, this may make one speculate

[74] William Flynn, "Hildegard (1098–1179) and the Virgin Martyrs of Cologne," in *The Cult of St Ursula and the 11,000 Virgins*, ed. Jane Cartwright (Wiltshire: University of Wales Press, 2016), 95.

[75] The best known hagiography of the Ursuline legend at the time would have been the *Regnante domino*, see María Eugenia Góngora, "Elisabeth von Schönau and the Story of St Ursula: Visionary Authority and the Cult of the Saints," in *Mulieres religiosae. Shaping Female Spiritual Authority in the Medieval and Early Modern Periods*, ed. Veerle Fraeters and Imke de Gier, Europa Sacra 12 (Turnhout: Brepols, 2014), 21.

[76] "Habebat quippe suspicionem de inventoribus sanctorum corporum, ne forte lucrandi causa titulos illos dolose conscribi fecissent," see Elisabeth Schoenaugiensis, *Visionen*, 124–5; translation from Elisabeth Schoenaugiensis, *Complete Works*, 215.

[77] "De his enim me silere non permittunt quidam bone opinionis viri, qui ad hec investiganda diutina me postulatione multum renitentem compulerunt." See Elisabeth Schoenaugiensis, *Visionen*, §1, 124; translation from Elisabeth Schoe-

about the political mechanisms operational behind the screens, and of the partisanship in Elisabeth's messages. This kind of "politics of mysticism" seems to have inspired later women religious, such as Elisabeth of Spalbeek (fl. 1246-1304), who employed her divine charism to also authenticate Ursuline relics, a service for which alliances and favours came in return.[78] In observing the unearthing of the Ursuline relics from the right bank of the Rhine, Gerlach too was a religious client interested in authenticating the relics in Cologne, and hoped to affirm his credibility by finding support in Elisabeth's visions.

It was no rarity in the twelfth century that the unmasking of forged relics would bring public discredit to those who endorsed them. Guibert of Nogent's polemic *De sanctis et eorum pigneribus* may serve as a fine example of this.[79] In light of the tension that came with the renewed suspicions, Gerlach's transference of two of the bodies may not have been a mere gift, but an inducement for which he expected favours in return. Much more than for any of Elisabeth's former visions, which —in Clark's words— revolved around "Elisabeth herself and what she experienced,"[80] one gains the impression of a 'paid job,' where Elisabeth's *Book of Revelations* functioned as a certificate. Aside from this word featuring literally in the text to describe what Gerlach is after (*cupiens certificare*),[81] the word choice of 'certificate' with its judicial connotation is more apt than one might presume. In recent years, both Góngora and Campbell stressed the judicial structure apparent within the last of Elisabeth's visions, which almost has the allure of a law suit.[82] The maiden Verena, one of the two martyrs' bodies appearing in Elisabeth's visions, is subjected to critical interrogation, and is to explain herself and the inconsistencies between two versions of the Ursuline legend. Elisabeth's questions to the apparition of Verena are to the point, and have the purpose of meticulous fact-checking. Who wrote the titles on the stones? Why, indeed, did men escort the eleven thousand virgins to Rome? Elisabeth goes as far

naugiensis, *Complete Works*, 213.

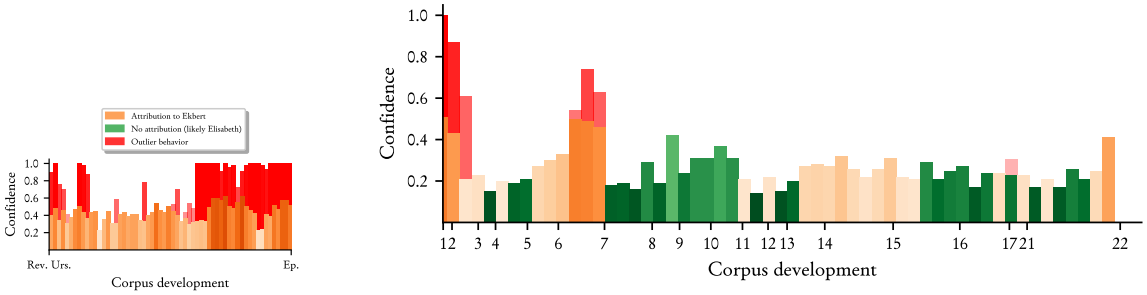
[78] See Philippe George, "A Saint-Trond, un import-export de reliques des Onze Mille Vierges dans la seconde moitié du XIIIe siècle," *Bulletin de la Société Royale Le Vieux-Liège* 12, nos. 252-3 (1991): 209–28; Jesse Njus, "The Politics of Mysticism: Elisabeth of Spalbeek in Context," *Church History* 77, no. 2 (2008): 308–9: "[...] In the mid-twelfth century, Elisabeth of Schönau had authenticated and identified a number of Ursuline relics through a series of visions that her brother recorded. Elisabeth of Spalbeek likely provided a similar service for the Ursuline relics of Abbot William of Ryckel, who was not only her relative but who also considered her his "spiritual daughter." While relic identification was not uncommon among holy women, Elisabeth not only authenticated and identified the relics but became a full participant in their exchange. Her guarantee of the relics' authenticity enabled their circulation among her associates, as well as among William's, and several of the recipients seem to have been acquainted primarily with her."

[79] The text's title translates to *On Saints and Their Relics*. In the text, Guibert criticizes the nearby located monastery of Saint-Médard which proclaimed to possess Christ's baby tooth. Guibert takes this as point of departure to express his annoyance over the preposterous types of miracular stories that abounded in his day, often fictitious so as to make earthly winnings. See Guibertus de Novigento, *De sanctis et eorum pigneribus*.

[80] Clark, *Elisabeth of Schönau*, 18.

[81] Elisabeth Schoenaugiensis, *Visionen*, 124.

[82] Mary Marshall Campbell, "Sanctity and Identity: The Authentication of the Ursuline Relics and Legal Discourse in Elisabeth von Schönau's *Liber Revelationum*," *The Journal of Medieval Religious Cultures* 38, no. 2 (2012): 159–92; Góngora, "Elisabeth von Schönau and the Story of St Ursula."




(a) Miniature rehearsal of fig. 5.5b (*Rev. Urs.*).

(b) Detail of Elisabeth of Schönau's letter corpus. The x-axis indicates the chapter number of the letter. Letters 5 and 6, both addressed to Gerlach of Deutz and indirectly preoccupied by the naming of the Colognian martyrs, are attributed to Ekbert and show irregular stylistic behaviour.

Figure 5.10: Details of fig. 5.4, exemplifying how two letters addressed to Gerlach of Deutz are likewise attributed to Ekbert of Schönau.

as to ask critical questions concerning the blatant anachronism in the popular legend: Attila the Hun could not have lived in the same period as Pope Cyriacus, one other martyr found amongst the Colognian corpses.[83] In the end, respondent Verena is able to satisfactorily answer her interrogator's questions, and the veracity of the Ursuline exhumation is guaranteed oracularly through Elisabeth's judgement.[84] Ultimately, her version of the Ursuline legend would become canonical.

Or, in light of the results as visualized in fig. 5.10, are we to say 'Ekbert's version'? Ekbert's traditional role of the interrogator (or prompter) in the Ursuline visions seems to have been extended towards taking on the role of respondent as well. All of the samples within *Rev. Urs.* are attributed confidently to him, which is unique in consideration of the neighbouring books. No other work is drawn this confidently to Ekbert in stylistic terms. Moreover, it has most of the internally anomalous sections of the corpus, which can be judged by the high, red peaks detected through the intrinsic plagiarism detection. And this is not all. Additional stylometric evidence further reinforces Ekbert's interference in other of Elisabeth's texts engaging with the Ursuline diggings. In some manuscript versions, three letters are appended at the ending of *Rev. Urs.*,[85] which can now be found in chapters 5, 6 and 17 of Elisabeth's letter corpus. All three letters are addressed to Gerlach of Deutz, and engage with the cult of the Cologne martyrs. Closer scrutiny of the letters in fig. 5.10 betrays that these

[83] See Elisabeth Schoenaugiensis, *Visionen*, §16, 132–3. Verena, then, summons the argument that not Attila, but some other prince by name of Julius was the leader of the Huns.

[84] "Ce type de visions fait preuve d'un nouveau côté, plus pragmatique, de sa prophétie: elle est consultée comme un oracle, un fait dont témoignent ses lettres ainsi que les très célèbres révélations ursulines dont il sera question plus loin," see Uta Kleine, "«Ce sont les mots que profère une langue nouvelle». Élisabeth de Schönau et le renouveau de la prophétie du XIIe siècle," in *Hagiographie et prophétie (VIe-XIIIe siècles)*, ed. Patrick Henriet, Klaus Herbers, and Hans-Christian Lehner, Micrologus Library 80 (Firenze: Sismel, 2017), 167.

[85] Elisabeth Schoenaugiensis, *Complete Works*, 297.

letters show conspicuous stylistic affinity to Ekbert. Especially letter 6 has a profile distinctively his and his alone.

Ekbert's dominant role in the wording of the Ursuline visions —and by extension in that of other correspondence to Gerlach of Deutz— has a significance on two levels. Firstly, it sheds a new light on the shared interests between Schönau and the ecclesial notables from Cologne, to whom both Hildelin (former abbot of Schönau) and Ekbert maintained close connections.[86] The intrusions up to the point of total stylistic control over the text testify of the sensitivities and urgencies at play in establishing the authenticity of the Ursuline remains. The second level of interest is of a cultural-historical nature. Elisabeth's initial reluctance to accept Gerlach's request, and the pressure she experienced from others' (forceful?) persuasion, receives a new overtone in light of the work's attribution to Ekbert. She does not speak of Ekbert alone, but of 'several men' (*quidam bone opinionis viri*) that would exert 'pressure' (e.g. *postulatione*, *compulerunt*) to write down her visions in spite of her reservations.[87] It will have been a work where every bit of phraseology mattered. Every choice of word counts when sensitive problems are to be tackled. Vital relations with important men were at stake, many of whom had direct gain in seeing their suggestions verified. A wide and interested audience awaited the work's publication, which will have been one of Elisabeth's first outside the Schönau convent walls.[88] In light of this context, Ekbert may have found Elisabeth's deficient grasp of Latin more problematic than ever.

## 5.4 Conclusive Remarks

This chapter has advocated computational stylistics as an alternative approach to explore the dynamics of cross-sex literary collaboration in Latin visionary literature (an idea which will largely extend into the next chapter on the *Vita Hildegardis*). By introducing the rolling *impostors* method, a novel method combining elements of intrinsic plagiarism detection and authorship verification, an attempt has been made to detect passages potentially preserving traces of Elisabeth of Schönau's writing style, be

[86] See Köster, "Das visionäre Werk Elisabeths von Schönau," 81; and Roth, *Visionen*, cxvi. Ekbert's former church — Saint Cassius in Bonn, where he had been deacon until his monastic vocation— fell under the jurisdiction of the Colognian diocese. He entertained close relations, for instance, with the man who would soon to become the diocese's archbishop: Rainald of Dassel (from 1159 until 1167). Ekbert might have been a fellow student of Rainald in Paris and dedicated his thirteen sermons against the Cathars to him. It was Rainald who offered Ekbert a promising career in the Church by offering him a position in Utrecht, to no avail, as Ekbert declined the offer in favour of Schönau. According to Köster, Ekbert also would have had relations to the former Colognian archbishop (and prince-elector) Arnold II of Wied († 1156). See Köster, "Leben, Persönlichkeit und visionäres Werk," 19.

[87] Elisabeth Schoenaugiensis, *Visionen*, §1, 123.

[88] The *Rev. Urs.* was to become Elisabeth's most important contemporary work, and the most widely read throughout the Middle Ages. See Clark, *Elisabeth of Schönau*, 37. "In fact, although Elisabeth's visions began in 1152 and the nuns of Schönau wrote down her descriptions of her experience, there is no evidence of a formal collection of visionary texts published outside the walls of Schönau until at least 1159," see Clark, *Elisabeth of Schönau*, 64.

it with the major remark that the intricate patchwork of (collaborative) styles in the visions heavily obstructed this attempt. Most importantly, this chapter has offered a bird's eye view on how a medieval collaborative composition process between the sexes can be characterized, both in its smallest nuances and in its larger contours.

Conspicuously enough, the foregoing experiments have suggested that a chronological trend in Elisabeth's visions can be discerned, revealing Ekbert's interferences over time as the sibling's writing partnership intensified. This in its own turn might present a degree by which we can expect Ekbert to have revised Elisabeth's former visions, which we surely know he must have done on several occasions.[89] Simultaneously, this visual rendering suggests that the collaboration between Elisabeth and Ekbert is a more nuanced and complex rope war between different parties (multiple collaborators might have been involved). When considering passages in detail, a number of discrete presuppositions in the state-of-the-art philological and historical research on the visions were called into question. The sections on the Cathars and the concluding vision of the third book betray little of Ekbert's influence, and —somewhat counter-intuitively— provide our best guesses at finding a style and mode of writing that is 'Elisabethan' or in which Elisabeth might have had the largest hand. If anything, what should definitely be questioned is if "Ekbert's erudition [...] at the end of the third *Liber visionum*" is indeed unilaterally Ekbert's doing.[90] Even though Ekbert's authorship is asserted in the text itself, something out of the ordinary happens in these particular passages which our stylistic analyses do not associate with Ekbert at all. A second suspect series of passages under scrutiny in this chapter were the revelations on Saint Ursula and the diggings around Cologne. This book heavily testifies of Ekbert's influence alone, which makes one reflect upon Elisabeth's and Ekbert's intensifying relations with the world at large and their role within the political network invested in the unearthing of the relics. Connected to this comes Ekbert's intensified concern over the contents of Elisabeth's writing, as his stylistic intrusions are more present than anywhere else in Elisabeth's works.

A final conclusion, then, is of a more general character and can be transferred to the debates in the chapters to follow: it is clear that not all text-internal cues on shifting authorship, or text-internal structures indicating which voice is speaking at what time, are very reliable. So, for instance, not because Ekbert is announced to start speaking at the end of *Vis.* III does his style surface more strongly. Equally so does the occurrence of Elisabeth's voice in the revelations on Saint Ursula not necessarily diminish Ekbert's extensive stylistic grasp over that particular text. It even appears that all the more alertness is advised when such explicit shifts of authorship are announced. The text-internal cues on authorship might not ever have arisen from a motivation to record a historical-material reality, but are in effect strategically implemented to grant

[89] Clark, *Elisabeth of Schönau*, 46.

[90] Coakley, *Women, Men and Spiritual Power*, 29.

individual portions of text the authority and authorship a medieval audience would have expected (and respected).

# 6

# Larger than Life? The *Vita Hildegardis*

## 6.1 Hildegard's Authorship and her Collaborators

### 6.1.1 *Die Echtheit*

The Rhenish visionary Hildegard of Bingen (1098–1179) has come a long way before becoming the celebrated twelfth-century female author she now is. Similarly to Elisabeth of Schönau, Hildegard composed her works in collaboration with male secretaries such as her provost and lifelong confidant Volmar of Disibodenberg († 1173), or her last secretary Guibert of Gembloux (1124–1214), and female scribes amongst whom her close companion Richardis of Stade.[1] Hildegard presented these collaborations as necessary to her readership due to her limited schooling and deficiency in speaking and writing Latin. In the nineteenth century, Wilhelm Preger took the presence of her male secretaries and self-devaluations as evidence that her entire *epistolarium* (both outgoing as incoming letters) was a *falsum*, and that only a man could have been capable of the intellectual accomplishments that speak from the remaining of Hildegard's works. Consequently, her entire oeuvre was ascribed to Theoderic of Echternach,[2] ironically enough a monk who might never even have met Hildegard in real life.[3]

Having for some time been denied authorship over her works, Hildegard has since Preger slowly but surely been regaining recognition for her work. Her revaluation was gradually set in motion with the pioneering work of Herwegen[4] and Liebeschütz[5] in the early twentieth century. The seminal work of the two Eibingen nuns Schrader

[1] On Hildegard's collaboration with secretaries, see especially the tripartite article of Hildephonse Herwegen, "Les collaborateurs de sainte Hildegarde," *Revue Bénédictine* 21, no. 1 (1904): 192–204; 302–316 (*Suite*); 381–403 (*Fin*).

[2] Preger, *Geschichte der deutschen Mystik*, 29–36.

[3] See p. 184 below, n. 41.

[4] See n. 1.

[5] Hans Liebeschütz, *Das allegorische Weltbild der Heiligen Hildegard von Bingen* (Leipzig: Teubner, 1930).

and Führkötter in 1956[6] established for good the authenticity of Hildegard's visionary trilogy (*Scivias*, *Liber vitae meritorum* and *Liber divinorum operum*). The second half of that same century also saw the arrival of critical editions by Fürhkötter, Carlevaris, Dronke and Derolez, Klaes and Van Acker,[7] and a series of scholarly publications by Dronke,[8] Newman[9] and Deploige[10] have been able to help us understand the 'Sibyl of the Rhine' as a unique author with an impressive breadth of intellectual accomplishments. She was not only a visionary writer, but also a musical composer, a playwright, a healer, a scientist, a writer of letters, an inventor of languages, and a theological thinker in her own right.

In a similar way as Elisabeth of Schönau, Hildegard's self-deprecations of her weaker sex (*paupercula forma*) and lack of education (*indocta*) are now taken to be of a topical nature, both to assert humility and to stress the miraculous and divine origin of her utterings. Even though she never directly cites any of her main sources, one must not underestimate the extent by which her illiteracy and unlearnedness are exaggerations with a strategic purpose.[11] In being prohibited to teach doctrine, Hildegard had to find a delicate balance between self-devaluation and self-authorization, between proclaiming her insufficiency whilst asserting her divinely inspired authority. The condition of male supervision —her closest secretary Volmar was the Rupertsberg's provost— undoubtedly shielded her against criticism from without, and gave Hildegard's visionary revelations a framework within the structure of the contemporaneous church.

Even though Hildegard's intellectual contribution to her own works is no longer contested, the contribution by her secretaries remains a matter of debate. The nature of the secretary question has, however, shifted. It does not wish to repeat Preger's far-fetched forgery theory, which has been refuted by an interlying century of scholarship, but arises from an interest in the historical reality of Hildegard's authorship and composition process. Even though Hildegard issued warnings that her secretaries were not to change the sense of her visions, and were to focus on formal aspects of the lan-

[6] Schrader and Führkötter, *Echtheit*.

[7] The editions of the three visionary works are Hildegardis Bingensis, *Scivias*, ed. Adelgundis Führkötter and Angela Carlevaris, *CC CM* 43 (Turnhout: Brepols, 1978), *Liber vite meritorum*, ed. Angela Carlevaris, *CC CM* 90 (Turnhout: Brepols, 1995); and *Liber divinorum operum*, ed. Albert Derolez and Peter Dronke, *CC CM* 92 (Turnhout: Brepols, 1996). The *epistolarium* is edited in three volumes in Hildegardis Bingensis, *Epistolarium*, ed. Lieven Van Acker and Monika Klaes-Hachmöller, *CC CM* 91–91B (Turnhout: Brepols, 1991–2001).

[8] Dronke, *Women Writers*.

[9] Barbara Newman, *Sister of Wisdom. St Hildegard's Theology of the Feminine* (Berkeley and Los Angeles: University of California Press, 1987).

[10] Deploige, *In nomine femineo indocta*.

[11] With regard to Hildegard's use of sources, Peter Dronke has noted "a sense [...] that verbal reminiscences have at times been deliberately covered over," see Peter Dronke, "The Allegorical World-Picture of Hildegard of Bingen: Revaluations and New Problems," in *Hildegard of Bingen: The Context of Her Thought and Art*, ed. Charles S.F. Burnett and Peter Dronke, Warburg Institute Colloquia 4 (London: Warburg Institute, 1998), 14.

guage such as grammar and spelling alone,[12] scholars such as Herwegen,[13] Derolez,[14] Kestemont, Moens and Deploige[15] have emphasized both by means of palaeographical and computational evidence that the influence of these secretaries remains —up until today— a fickle point of Hildegard scholarship indeed. The extensive corrections by secretaries in the Ghent MS 241 apograph of the *Liber divinorum operum*, for instance, may make one wonder to what extent the redaction and revision process by Hildegard's secretaries profoundly changed the final outlook of the visionary's works.[16] Kestemont, Moens and Deploige have provided stylometric evidence to show Guibert of Gembloux's stylistic influence on two suspect visions, *Visio ad Guibertum missa* and *De excellentia Sancti Martini*,[17] and a somewhat more subtle yet noticeable stylistic presence in Hildegard's letters written after Volmar's death in 1173 (likely to have been revised by Guibert).[18]

### 6.1.2 Hildegard's Image: The Riesencodex

In this chapter, we address the notion that Hildegard's secretaries played a vital role in shaping her image and authority, both during her life and after her death. It posits that the aspect of collective creativity is fundamental in understanding both Hildegard as figure and as author, *without* seeking to undermine Hildegard as figure and author. Hildegard's collaboration with secretaries did not unilaterally constitute her fabrication (as Preger argued earlier), neither did it unambiguously signify her suppression by a male patriarchy (i.e. the willful alteration of her words by male secretaries). Both of these poles give no satisfactory explanation of the dynamics at play, but bypass a much more complex tension field of constant negotiation in which mutual interests are at stake. As Johnson has noted, one should not forget that Hildegard's authority as writer benefited greatly from the endorsement and encirclement by male clerics, an effect which she cannot but have been aware of, and therefore incorporated into her texts.[19] This kind of subtle interplay between asserting her authority and having her authority asserted through involvement of others is always present. For instance, Hildegard grants Volmar a central role in having launched her writing career at the age of forty-two, but on the other hand fails to ever mention his name.[20] Hildegard

[12] Ferrante, "*Scribe quae vides et audis*," 103.

[13] See n. 1.

[14] Derolez, "Deux notes."

[15] Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century."

[16] See Derolez's introduction to the edition of the *Liber divinorum operum*, lxxxixff., referenced earlier in n. 7.

[17] Both edited in Hildegardis Bingensis, *Visiones Hildegardis a Guiberto Gemblacensi exaratae*, in *Hildegardis Bingensis opera minora* II, ed. Jeroen Deploige and Sara Moens, *CC CM* 226A (Turnhout: Brepols, 2016), 135–236.

[18] Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 213–8.

[19] "Hildegard thus underlines her authority as both author and seer, an authority manifested through Volmar," see Johnson, "The Trope of the Scribe," 823.

[20] Only Guibert of Gembloux does in his *epistolarium*. See the introduction to Hildegardis Bingensis, *Visiones Hildegardis*, 141.

(and/or her entourage?) perceptibly sought for mechanisms by which to authorize her visionary writings by involving onlookers and alliances that recognized the divine origins of her extraordinary gift. Such attestations, however, are better not always accredited at face value. Volmar is known to have concocted a false letter from Pope Eugene III, which was meant to present a prestigious first letter heading her *epistolarium*.[21]

Hildegard's authorship can be thought of as a project, and the Rupertsberg community had stakes in its success long after she was gone. Her dossier was intensely prepared in the final years of her life, the period from which most contemporary Rupertsberg manuscripts survive. The Wiesbaden Riesencodex, a monumental codex containing Hildegard's *opera omnia* of all authorized versions of her visionary writings (except for her *Liber subtilitatum diversarum naturarum creaturarum*, or *Physica* and *Cause et Cure*), has been argued to have been issued under Hildegard's command, but was completed after her death.[22] It was Hildegard's persona rather than her historical person that was documented and 'constructed' in the Riesencodex.

Preger's choice in designating a male forger for Hildegard's entire oeuvre fell on Theoderic of Echternach. This was not a coincidental choice. Theoderic was the end redactor of the *Vita Hildegardis*, the "Life of Hildegard." In Preger's eyes, being the orchestrator of Hildegard's life must have only been a few steps removed from being its overall fabricator. The *Vita* has been a source of fascination to scholars not only because it gives an account of Hildegard's life, but also because it contains direct citations of 'autobiographical' fragments allegedly written by the visionary herself. The editor of the *Vita*, Monika Klaes, seemed to have been unconvinced by Schrader, Führkötter[23] and Dronke's[24] convictions that these passages contain no interpolations and are authentically Hildegardian.[25] Quite recently, Van Engen confessed doubts concerning their authenticity as well: "To what degree they, even if of authentic origin, underwent redaction lies beyond our ken."[26] Taking into consideration that

[21] Van Engen, "Letters and the Public Persona of Hildegard," 380. The spurious letter is edited by Van Acker in Hildegardis Bingensis, *Epistolarium Pars Tertia CCLI–CCCXC*, 173.

[22] MS Wiesbaden, Hessische Landesbibliothek, 2. The bulk of the codex must have been finished during Hildegard's lifetime, before 1179, and most certainly is a product of the Rupertsberg scriptorium. Therefore, the manuscript has a most intricate life cycle. Five to six hands have been discerned in the codex, and it is likely that some of the scribes responsible for the Riesencodex's transmission have worked simultaneously. The end redaction is now generally attributed to Guibert of Gembloux, a Brabantine monk who, as we will discuss further down, played an important role in the collecting and completing of Hildegard's oeuvre at the end of her life. The question has often been raised if the Riesencodex is a unique occasion in which a twelfth-century author personally supervised the process of redaction and compilation of an *opera omnia* in a single volume. For the state-of-the-art codicological findings on the Riesencodex, see Michael Embach, *Die Schriften Hildegards von Bingen: Studien zu ihrer Überlieferung und Rezeption im Mittelalter und in der Frühen Neuzeit*, Erudiri sapientia: Studien zum Mittelalter und zu seiner Rezeptionsgeschichte 4 (Berlin: Akademie Verlag, 2003), 36–65.

[23] "Der einfache Stil dieser autobiographischen Einfügungen trägt durchaus den Charakter der Schriften Hildegards, und es ist nicht ersichtlich, warum die Echtheit dieser Texte anzuzweifeln wäre," see Schrader and Führkötter, *Echtheit*, 14.

[24] Dronke, *Women Writers*, 144.

[25] Klaes, *Vita Hildegardis*, 113*–114*.

[26] John Van Engen, "Authorship, Authority, and Authorization: The Cases of Abbot Bernard of Clairvaux and Abbes Hildegard of Bingen," *Lectio* 4 (2016): 339.

the *Vita* was interpolated in Hildegard's Riesencodex posthumously,[27] that it went through the hands of at least three biographers (perhaps four or more), and that it had the intention of promulgating Hildegard's persona with lasting impact, one will see why due concern is warranted. Is this indeed Hildegard's *Life*, or is this a larger-than-life literary construction, and how was she herself participant to its composition?

## 6.2 Composition History

When speaking of Hildegard's *Vita*, we are hardly speaking of a 'single' text. Multiple *Vitae Hildegardis* have co-existed and circulated. A few of them are finished and transmitted, others are recoverable only to a certain extent, and there are versions irrecoverably lost. We currently have two more or less 'final' versions of what is officially known as the *Vita Hildegardis* at our disposal.[28] The first version, widely regarded as 'canonical,' is the integral text taken up in Monika Klaes's edition.[29] The second version is a revision of this canonical version by Guibert of Gembloux, which is recoverable through a critical apparatus and an appendix in the aforementioned edition. The complex textual history of Hildegard's *Vita*, which is elaborately discussed in Klaes's lengthy introduction with a number of pages twice as long as the actual text, teaches us that its composition was accompanied by many difficulties. Despite best intentions, the project was on the verge of deferral due to a chain of unfortunate circumstances and unexpected deaths before the *Vita*'s completion, not the least Hildegard's own death in 1179. Consequently, as will be further discussed, here is a text with an intricate timeline, and an archetypal example of collaborative authorship. Five authors were (at least partially) involved, the first of which allegedly Hildegard herself. The *Vita* contains so-called autobiographical fragments and snippets of visionary material, apparently dictated in the first person by Hildegard herself, which cannot be found elsewhere in Hildegard's oeuvre. The remaining (co-)authors are her secretaries and biographers, in —more or less— chronological order of contribution: Volmar of Disibodenberg, Godfrey of Disibodenberg, Guibert of Gembloux

[27] Not the Riesencodex, but an autograph by Theoderic was Klaes's choice for main manuscript. See Klaes, *Vita Hildegardis*, 157*ff. Also see Albert Derolez, "The Manuscript Transmission of Hildegard of Bingen's Writings: The State of the Problem," in *Hildegard of Bingen: The Context of Her Thought and Art*, ed. Charles S.F. Burnett and Peter Dronke, Warburg Institute Colloquia 4 (London: Warburg Institute, 1998), 23.

[28] I leave out of regard for now the anonymous *Octo lectiones in festo Sanctae Hildegardis legendae* and the abbreviated *Vita* (*abbreviata Traiectensis*) by Guibert of Gembloux, both of which are derivations from the 'official' *Vita* that will be discussed here. Both texts are likewise edited in Godefridus Sancti Disibodi and Theodericus Epternacensis, *Vita Hildegardis*, respectively at 75–80 and 83–8. I also pass over the *Acta inquisitionis*, sent to Pope Gregory XI in 1233 to Rome by three clerics in Mainz for approval of Hildegard's canonization. Included was Theoderic's *Vita Hildegardis* and exemplars of her work. See Petro Bruder, ed., *Acta inquisitionis de virtutibus et miraculis S. Hildegardis, Magistrae sororum ord. S. Benedicti in monte S. Ruperti juxta bingium ad rhenum. Ex originali archetypo transcripsit notisque illustravit*, *Analecta Bollandiana* 2 (1883), 116–29. See George Ferzoco, "The Canonization and Doctorization of Hildegard of Bingen," in *A Companion to Hildegard of Bingen*, ed. Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, Brill's Companions to the Christian Tradition 45 (Leiden: Brill, 2014), 306.

[29] Godefridus Sancti Disibodi and Theodericus Epternacensis, *Vita Hildegardis*.

and finally Theoderic of Echternach.

### 6.2.1 Hildegard, Volmar and Godfrey of Disibodenberg

It has generally been accepted that the collecting of materials and the drafting of early versions of the *Vita* had already begun under Hildegard's direction.[30] In this she was assisted by her secretaries, principally Volmar and —after the latter's death in 1173— by Godfrey of Disibodenberg, who arrived in the Rupertsberg shortly after in 1174. Around 1175/6, only a year and a half later, also Godfrey would come to die.[31] By the time of the latter's death, a partial *Vita* for Hildegard had been composed, a 'booklet' or *libellus* which is believed to be transmitted as the first book of the complete *Vita Hildegardis*. Whether or not the accounts collected in the first book of the *Vita* are indeed an integral copy of Godfrey's original booklet is unknown: other secretaries of Hildegard and her succeeding biographers will have had ample opportunity to revise the text.[32]

### 6.2.2 Guibert of Gembloux

The first of Godfrey's successors was Hildegard's last secretary Guibert of Gembloux (c. 1124–1214), a Brabantine monk who came to her aid from 1177 onwards and would assist Hildegard until she passed away two years later. During his time at the Rupertsberg, the Colognian archbishop Philip I of Heinsberg appears to have instigated in Guibert the intention of writing a *Vita* of his own.[33] Guibert hesitated to obey Philip's request during Hildegard's lifetime, fearful of being found sycophantic by Hildegard, and it was only after her death in September 1179 that —in his research of materials— he bumped into useful sources to facilitate the task. From the description of his findings, one can suspect that he had come across autobiographical memoirs of Hildegard and Volmar, and a *libellus* that might well correspond to Godfrey's first *Vita*.[34] Therefore, some time between Hildegard's death and Guib-

[30] Concerning Hildegard's role in the *Vita*, Newman has emphasized twice that she was likely an early orchestrator, see Newman, "Three-Part Invention"; and Barbara Newman, "Hildegard and Her Hagiographers. The Remaking of Female Sainthood," in *Gendered Voices. Medieval Saints and Their Interpreters*, ed. Catherine M. Mooney, The Middle Ages Series (Philadelphia: University of Pennsylvania Press, 1999), whether or not the aim was to strive for Hildegard's canonization and the *Vita* as crowning piece of the *Riesencodex opera omnia*, is difficult to ascertain, see Klaes, *Vita Hildegardis*, 78*.

[31] See Albert Derolez's introduction to Guibertus Gemblacensis, *Epistolae*, vi–vii.

[32] As stated in Klaes's introduction to the edition, see *Vita Hildegardis*, 91*. According to Newman, Godfrey's first book might originally have been a first person singular memoir dictated by Hildegard. This might be possible, although Newman admits that there is no real evidence for this, see Newman, "Hildegard and Her Hagiographers," 17 .

[33] This becomes clear from Guibert's letter to Philip I. See Guibertus Gemblacensis, *Epistolae*, Ep. 15, 210–15; the letter is difficult to date, but Klaes has shown that the letter's contents suggest that Guibert wrote it in 1180, at which time he was still in Bingen, and had just caught news of his being recalled to Gembloux. See Klaes, *Vita Hildegardis*, 30*–31*.

[34] All of this is described in a letter to Philip I, see n. 33. Because the *termini post* and *ante quem* for the dating of the letter lie in between Hildegard's death in 1179 and that of Philip I in 1191, Klaes demonstrates in her introduction to the *Vita* that there might well be some ambiguity concerning the *libellus* referred to, see Klaes, *Vita Hildegardis*, 30*.

ert's departure from the Rupertsberg, progress for a new *Vita* appears to have been well under way. These plans were disturbed when —after Easter 1180— Guibert was forced to return from the Rupertsberg on appeal of the abbot of Gembloux.[35] It appears that this event left Guibert unable to finish the work he had started on his *Vita*, although a fragment of his efforts at the time is presumed to have survived as an attachment to a letter addressed to his fellow monk Bovo of Gembloux.[36]

What should be emphasized here, is that Guibert's last year in Bingen (1179–80), which revolved entirely around the collecting of the prime sources of the *Vita*, left him with the ample opportunity to rework and revise Hildegard's materials —particularly the 'autobiographical' fragments—, and equally so the preparatory versions by Volmar and Godfrey, both of which are unknown in their original state.[37] Former research has already pointed out that the large role Guibert played in compiling and editing Hildegard's works cannot be underestimated, an activity which set out after his arrival in 1177, in the last two years of the visionary's life. Guibert supervised the scriptorium's activities at a time when Hildegard's *epistolarium* and the Riesencodex, both discussed earlier on p. 179, were in the final stages of completion.[38] As has recently been pointed out by Kestemont, Moens and Deploige, Guibert appears to have been granted —or appears to have taken— unprecedented liberties in editing and revising Hildegard's works, as is asserted in a letter from Hildegard which carries Guibert's style completely.[39] Considering Guibert's mark on Hildegard's works in the final stages of her life, and considering that he did not flinch from making alterations to the visionary's wording, one may foster justified wariness concerning his early involvement in the composition of the *Vita* as well.[40]

Either Guibert's *libellus* refers to Godfrey's first book, or else it corresponds to the redacted version by Theoderic. Klaes's arguments of the dating of the letter to Philip I, however, already expounded upon in n. 33, strongly suggest that Guibert is speaking in 1180, which was at a time when Theoderic of Echternach was yet to arrive in Bingen. Therefore, the *libellus* referred to is likely Godfrey's. The impression that Guibert collected sources and wrote a fragmentary *Vita* while still at Bingen the year after Hildegard's death, is further reinforced by another one of his letters, namely that to monk Bovo, for which I refer to n. 36.

[35] Sara Moens, "De horizonten van Guibertus van Gembloers (ca. 1124–1214). De wereld van een benedictijns briefschrijver in tijden van een verschuivend religieus landschap" (PhD diss., 2014), 74.

[36] The letter itself purports to have been written in 1177, but as Derolez argues it was —in the state that we have it— finished at the end of 1179 and the beginning of 1180. See Guibertus Gemblacensis, *Epistolae*, Ep. 38, 367–79; another strong argument for dating the fragment at a time when Guibert was still in Bingen, is a remark within the fragment itself stating "Hoc non solum cum adhuc in monte sancti Disibodi cum paucioribus degeret obseruauit, sed maxime hic, hoc est in Binguia," see ibid., ll. 330–2, 376. Guibert's *Vita* fragment breaks off mid-sentence in the best conserved manuscript of his letters, namely MS 5527–34, Royal Library, Brussels, often found abbreviated as $G^2$. Klaes presumes that it was originally longer, see *Vita Hildegardis*, 42*–43*.

[37] Emphasized in Klaes, *Vita Hildegardis*, 58*.

[38] Van Acker was the first to suggest that Guibert was responsible for Hildegard's *epistolarium* as it is presented in the Riesencodex, see Lieven Van Acker, "Der Briefwechsel der heiligen Hildegard von Bingen Vorbemerkungen zu einer kritischen Edition (*Fortsetzung*)," *Revue Bénédictine* 99, nos. 1–2 (1989): 129–34.

[39] Revised to the extent that Hildegard's style is hardly still recognizable. See Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 202–4. Also see p. 179 earlier.

[40] A point made by Klaes, *Vita Hildegardis*, 113*.

### 6.2.3 Theoderic of Echternach

Volmar's and Godfrey's deaths, and Guibert's commitment to new priorities on his path, left the late Hildegard yet without a *Vita*. On instigation of abbots Ludwig and Godfrey of Saint-Eucharius, the task consequently fell onto Theoderic of Echternach, an unlikely candidate, since the latter possibly never met Hildegard personally.[41] Interestingly, at this decisive juncture of the *Vita*'s gradual materialization, a branching off between Guibert's and Theoderic's versions seems to have taken place. Guibert lost track of Theoderic's progress on the *Vita*, whereas Theoderic shows no familiarity with Guibert's fragment, which would nevertheless have existed at the time. Still, there are parallels between both writers' *Vitae*, which indicates their dependence on the same pool of consulted source materials that was first collected by Guibert.[42] From the listing of sources in his preface, Theoderic indeed appears to have consulted the same sources for his *Vita* as Guibert, namely Godfrey's unfinished *libellus* and snippets of Hildegard's visions.[43] This means, as was raised earlier, that Theoderic might have used source materials heavily revised by Guibert. Theoderic's role then, was that of editor-in-chief, a role corresponding to a kind of narrator or commentator, tying together the seemingly unrelated bits and pieces that had coincidentally fallen into his hands. The general structure of the *Vita*, then, and the purported authors of its constituents, is the following:

| **Author** | **Title** (or ***incipit***) | **Ed.** (M. Klaes) |
|---|---|---|
| Theoderic of Echternach | *Prologus in vitam* | 3–4. |
| | *—capitula—* | 5. |
| Godfrey of Disibodenberg | I. *Liber primus* (*libellus*) | 6–16. |
| Theoderic of Echternach | *Prologus in librum secundum* | 17–8. |
| | *—capitula—* | 19. |
| | II. *Liber secundus* | 20–45. |
| | *Prologus in librum tercium* | 46. |
| | *—capitula—* | 47–8. |
| | III. *Liber tercius* (*De miraculis*) | 49–71. |

That Silvas translates *textus*, which is the word Theoderic uses in his prologue

[41] Klaes refuted the suggestions that Theoderic was one of the assistants who finished the *Liber divinorum operum* after Volmar's death in *Vita Hildegardis*, 60*–61*. She draws the conclusion that his occupations in Echternach as chronicler would have been too occupying for this hypothesis to hold true, see ibid., 77*.

[42] This is most conspicously clear from both *Vitae*'s usage of a fragment from Hildegard's letter to Guibert, the *De modo visionis sue*. See Klaes, *Vita Hildegardis*, 48*. The complete letter is edited as Ep. $103^{R}$ in Hildegardis Bingensis, *Epistolarium Pars Secunda XCI–CCL*$^{R}$, ed. Lieven Van Acker, *CC CM* 91A (Turnhout: Brepols, 1993), 258–65. The specific extract quoted in the *Vita* is found in ll. 54–75, pp. 260–1.

[43] "[...] Accepi, ut post Godefridum, uirum ingenio clarum, uitam sancte ac Deo dilecte Hildegardis uirginis, quam illo honesto stilo inchoauit, sed non perfecit, in ordinem colligerem et quasi odoriferis floribus serta contexens uisiones eius gestis suis insertas sub diuisione librorum in unius corporis formam redigerem," found in the *prologus* of Godefridus Sancti Disibodi and Theodericus Epternacensis, *Vita Hildegardis*, ll. 5–10, 3.

when referring to the *Vita*, as "tapestry" instead of simply "text," is significant.[44] The *Vita* truly is an interwoven assembly of impressions, gathered from different sources. Theoderic combined 1. Godfrey's *libellus*, 2. a number of 'autobiographical,' memoir-like visions, and 3. a number of performed miracles, again interspersed with Hildegard's visions. In assembling the *Vita* according to this schema, the monk asserted to have changed very little to their contents.[45] Yet, we must be careful to take Theoderic's word for this. Klaes, basing herself on a study of his style in his chronicle of Echternach, raised suspicions that some passages in the first book (Godfrey's *libellus*) betray his interventions.[46] She even argued that whole chapters within this same book, namely 8 and 9, are uniquely Theoderic's additions.[47] Those passages that are intact from Theoderic's adjustments generally exhibit a soberer character, and a simpler syntax, features that might have been typical for Godfrey's writing, but of whom we know very little and possess no written documents. Klaes is somewhat more hesitant as to the possibility of Theoderic's alterations to Hildegard's texts. Theoderic seemed too intimidated by the density of her visions to dare make any profound changes to them.[48]

### 6.2.4 Guibert's Revisions

Only in 1208/9, near the end of his life and some thirty years after Hildegard's death, Guibert of Gembloux acquaints himself with the *Vita* as redacted by Theoderic. Having in the meantime become monk in Florennes,[49] he asks for Hildegard's parents' names in a letter exchange with Godfrey of Saint-Eucharius, because he is writing a "little something" about the *magistra*'s life.[50] The need for refreshing his memory on Hildegard's biographical details indicates Guibert's renewed intentions of finishing the *Vita* left incomplete when leaving Bingen, and which, indeed, makes no mention of Hildegard's parents names.[51] In response, abbot Godfrey sends back Theoderic's *Vita*, and simultaneously solicits Guibert's corrections and additions, because still much is missing in Theoderic's impersonal account of the prophetess. Guibert answered God-

[44] See Godefridus Sancti Disibodi and Theodericus Epternacensis, *The Life of Hildegard*, in *Jutta and Hildegard: The Biographical Sources*, ed. and trans. Anna Silvas, Medieval Women: Texts and Contexts 1 (Turnhout: Brepols, 1998), 135. For the original Latin, see Godefridus Sancti Disibodi and Theodericus Epternacensis, *Vita Hildegardis*, 3, l. 19.

[45] Especially for Hildegard's autobiographical fragments in books 2 and 3, Theoderic insists that he has left them unaltered: "in descriptione uisionum eius nullatenus mutilaretur," see the prologue to the second book, in Godefridus Sancti Disibodi and Theodericus Epternacensis, *Vita Hildegardis*, II Prol., ll. 31–2, 18; he argues much the same for Godfrey's *libellus*, "nullam sue dispositionis patiatur iacturam," in ibid., I Prol., ll. 16–17, p.3.

[46] For instance the aretalogy of chapter 2, see Klaes, *Vita Hildegardis*, 95*–97*.

[47] Klaes, *Vita Hildegardis*, 92*–94*.

[48] Klaes, *Vita Hildegardis*, 111*.

[49] Guibert had been abbot of Florennes from c. 1188/9, and abbot of Gembloux from c. 1193/4 to 1204. After this, Guibert denounced abbacy, and became an ordinary monk in Florennes, see Moens, "De horizonten," 77–9.

[50] "Scripsi enim de illa aliquid, ubi libenter ea inseruissem, si recolere potuissem," see Guibertus Gemblacensis, *Epistolae*, Ep. 40, ll. 20–2, 385.

[51] As indicated in n. 36, the fragmentary *Vita* is appended to a letter to Bovo of Gembloux.

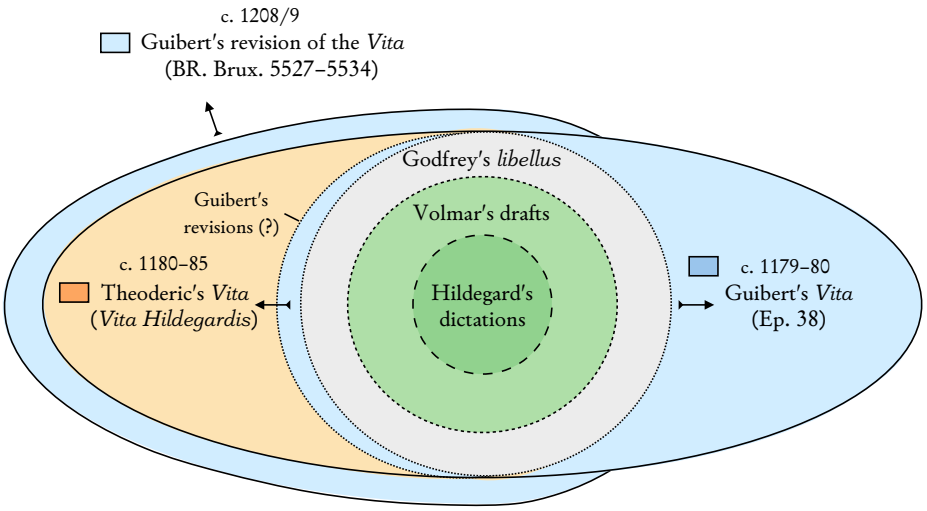



Figure 6.1: Schema of the composition stages of the *Vita Hildegardis*, visualizing its layered character and composite authorship. Full lines indicate extant works, dotted lines indicate lost works. Yellow indicates 'redacted by Theoderic of Echternach,' blue 'by Guibert of Gembloux.' The concentric circles at the core represent the original source material of the *Vita*, used by both Guibert of Gembloux in 1179/80 and Theoderic up until the mid-80s. These are supposedly Hildegard's 'original' memoirs, drafted or perhaps once transcribed by Volmar and/or Godfrey in a *libellus*. In the peripheries we see the two *Vitae* by Theoderic and by Guibert, reliant on the central sources yet independently composed. The blue layer separating the source material from Theoderic's *Vita* indicates the potential interferences made by Guibert, who collected the source material. Correspondingly, the blue layer wrapped around Theoderic's *Vita* on the outer edges indicates the revisions which are transmitted in MSS 5527–5534, and appended to Klaes's edition.

frey's request by stating that he found no fault in the work sent to him, and that his own fragmentary *Vita* could impossibly surpass a work of such great accomplishment. As far as we know, Guibert kept his word, and never completed a *Vita* of his own. But his reluctance to contribute to Theoderic's version, which features so strongly in his letter to Godfrey, appears to have been false modesty. A heavily stylistically altered version of Theoderic's *Vita* survives in both manuscripts of Guibert's letter collection. Guibert's interferences extend well into Hildegard's autobiographical fragments in book 2 and 3, which warrants concerns over the monk's general habit to revise Hildegard's texts. Especially small function words were Guibert's favourite target, pronouns such as *hic*, *ille* and *iste*, comparative conjunctions such as *quemadmodum* and *uelut*, etc.[52] Meanwhile, Guibert also appears to have been sensitive to Theoderic's interventions in Godfrey's *libellus*, as becomes clear from his restructuring of the first chapter.[53]

### 6.2.5 The *Vita*'s Voices: A Survey of Candidate Authors

Fig. 6.1 roughly sketches the composition history of the *Vita* as outlined in the previous section, and summarizes the (potential) zones of overlap between the different text versions. Especially the autobiographical passages as contained within the *Vita* have drawn a great deal of interest in Hildegard scholarship. Their existence prior to Theoderic's integration raises a few compelling questions as to their intended form and aim. Do they indicate, for instance, that already during her life, Hildegard was consciously constructing her self-image, with the objective of canonization in mind?[54] In relation to this question, one can wonder what could have been the original connection between the individual fragments transmitted in the *Vita* as we have it, and whether or not they derive from an originally integral 'autobiographical' *Vita*. Interestingly, even Godfrey's *libellus*, a third-person account of Hildegard's life, has been hypothesized by Newman to originally have been a first-person attestation by Hildegard, a true memoir, which Godfrey then rewrote from a different focalization point.[55] The question is if that would mean that Hildegard's original style is to a certain extent recoverable from Godfrey's transcription. On the other hand, we have the interference of a considerable number of male co-writers that had ample opportunity to rewrite and overwrite Hildegard's original text. Godfrey and Volmar present the first filters through which her signal passed. Guibert of Gembloux, then, whom we know was capable of altering Hildegard's style and did not flinch from doing so,[56] could have revised the material whilst collecting it, after which Theoderic of Echternach selected portions from it and possibly again revised all materials according to his own principles. The aim of the subsequent paragraphs is to shed more light on the extent to which these collaborators' treatment of Hildegard's text included the compromising of her language. More generally, our findings consequently invite reflection on Hildegard's authorship, and on the extent by which her involvement in the *Vita*'s composition is reflected stylistically in the text.

Assessing the authorship(s) of the *Vita* is a complex matter, because not all of the candidates have left independent writings of their own which facilitate a direct basis of comparison with a sample from the *Vita*. This is true for secretaries Volmar and Godfrey of Disibodenberg, or other potentially involved assistants of whose contributions we might simply be unaware. For the remaining authors of whom we do have the ability to assemble a background corpus —Hildegard, Guibert and Theoderic—, is-

[52] Klaes, *Vita Hildegardis*, 155*.

[53] Klaes, *Vita Hildegardis*, 95*.

[54] Newman seems confident that canonization was the aim of the whole project, but Klaes is somewhat more hesitant. See n. 30 on this topic.

[55] Newman, "Hildegard and Her Hagiographers," 17–8.

[56] Through function word analyses, Kestemont et al. were able to show that two of Hildegard's latest visions, the *Visio de sancto Martino* and *Visio ad Guibertum missa*, exhibit all the characteristics of Guibert of Gembloux's style, see Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century."

sues of reliability and incompatibility with the *Vita*'s genre and style are at stake. The background corpus is therefore small and comes with its proper insecurities. Guibert of Gembloux's *Epistolae* (± 124,500 *w*) and his very short *De combustione* (± 1,000 *w*) were included,[57] as were Theoderic of Echternach's chronicles.[58] For Hildegard of Bingen, her *Vitae* of Saint Disibod (± 7,500 *w*) and Saint Rupert (± 4,200 *w*) were integrated in the corpus,[59] which only seemed reasonable considering that these works best represent her handling of the hagio- and biographical genre to which the *Vita Hildegardis* belongs. Nevertheless, these text's shortness required the involvement of her visionary treatises as well: *Scivias*, *Liber vitae meritorum* and *Liber divinorum operum*. It is on the basis of these works that a training corpus was assembled that could best represent the stylistic profiles of these three important candidates. Hildegard's texts, which should provide the 'gold standard' of her style, can obviously not be strictly separated from the potential influences of Volmar. Theoderic's chronicles, the *Chronicon Epternacense* and the *Libellus de libertate Epternacensi propugnata*, contain genre-specific qualities that might destabilize a firm basis for comparison.[60]

## 6.3 Experimental Set-up

Having taken into consideration the training corpus at our disposal and the *Vita*'s specific difficulties, it becomes clear that the authorship problem of the *Vita Hildegardis* can be situated somewhere between traditional attribution ("amongst a set of candidates, choose the most suited one") and authorship verification ("is there a suited candidate amongst all candidates, yes or no"). We have traditional attribution, on the one hand, because for three of the most important candidates, Hildegard, Guibert and Theoderic, we have samples of their style. This is a 'closed' setting. Verification, on the other, because for potential other candidates —Volmar and Godfrey, to name but two— no such profile can be drawn up: an 'open' game.

As in the previous chapter, we chose a rolling method identical to that explained on p. 155. The *Vita* is a short text (± 15,300 *w*), potentially containing the influence of multiple authors. Note that the *Capitula* (after each prologue) were removed, for they cannot provide a sample of style as such. Instead of dividing up the text into discrete samples, we slide over the text by taking it in by 500 words at a time, and gradually proceeding onto the next sample by a step size of 100 words. For each of

[57] See Guibertus Gemblacensis, *Epistolae*; and Guibertus Gemblacensis, *De combustione monasterii Gemblacensis*, ed. Georg Heinrich Pertz, *MGH Scriptores* 8 (Hannover, 1848), 563–4.

[58] For reference to Theoderic's chronicles, see n. 60.

[59] Hildegardis Bingensis, *Vita Sancti Disibodi episcopi*, in *Hildegardis Bingensis opera minora* II, ed. Christopher P. Evans, *CC CM* 226A (Turnhout: Brepols, 2016), 59–87; Hildegardis Bingensis, *Vita Sancti Ruperti confessoris*, in *Hildegardis Bingensis opera minora* II, ed. Christopher P. Evans, *CC CM* 226A (Turnhout: Brepols, 2016), 91–108.

[60] See Theodericus Epternacensis, *Chronicon Epternacense*, in *MGH Scriptores* 23, ed. Ludwig Weiland (Hannover, 1874), 38–64; and Theodericus Epternacensis, *Libellus de libertate Epternacensi propugnata*, in *MGH Scriptores* 23, ed. Ludwig Weiland (Hannover, 1874), 64–72.

the 140 partially overlapping samples an assessment of authorship is made by applying two methods:

- The rolling *impostors* method[61] As usual, the impostors were drawn mostly from the benchmark corpus (A.I.1, p. 304).
- SVM (support vector machine)[62]

### 6.3.1 Rolling *impostors*

As has been explained a few times by now,[63] the *impostors* method's results indicate the $k$ times out of 100 that a candidate was chosen against a pool of potential impostors (the impostors in the benchmark corpus, see p. 304). As opposed to the experimentation with Ekbert in the previous chapter, the $\sigma^*$ threshold of 0.22 was in fact quite stable for distinguishing between Hildegard, Guibert and Theoderic, yielding quite promising figures despite the short sample length (see table 6.1). The PR-curve, given in fig 6.1, indicates at what point the number of $k$ attributions becomes truly reliable in a training phase.

| | dev set | test set |
|---|---|---|
| **accuracy** | 0.66 | 0.62 |
| **precision** | 0.88 | 0.89 |
| **recall** | 0.39 | 0.27 |
| **f1** | 0.54 | 0.41 |

Table 6.1: Evaluation metrics for training the *impostors* method on Theoderic of Echternach, Guibert of Gembloux and Hildegard of Bingen.

### 6.3.2 Support vector machine (SVM)

As previously explained in this thesis (p. 141), we trained multiple classifiers with different parameters.[64] This enabled us to rely on the prediction made by the majority of classifiers (indicated by a confidence score) instead of selecting solely the best-ranking one. The average performance ranged from 95.81% (worst) to 99.63% (best).[65] Fig. 6.3 yields a two-dimensional PCA plot visualizing how the best-performing SVM

[61] For the *impostors* method, see Koppel and Winter, "Determining."

[62] For a more detailed explanation of support vector machines (SVM), see chap. 3 on pp. 94ff., and this thesis's appendix on pp. 336ff. For its application to computational stylistics specifically, see Diederich et al., "Authorship Attribution with Support Vector Machines."

[63] The *impostors* method was explained in full in 3.4.3 (p. 117).

[64] These settings were the same as on p. 141.

[65] These percentages were based on a mean over accuracy, precision and recall, the concepts of which are explained on p. 342.

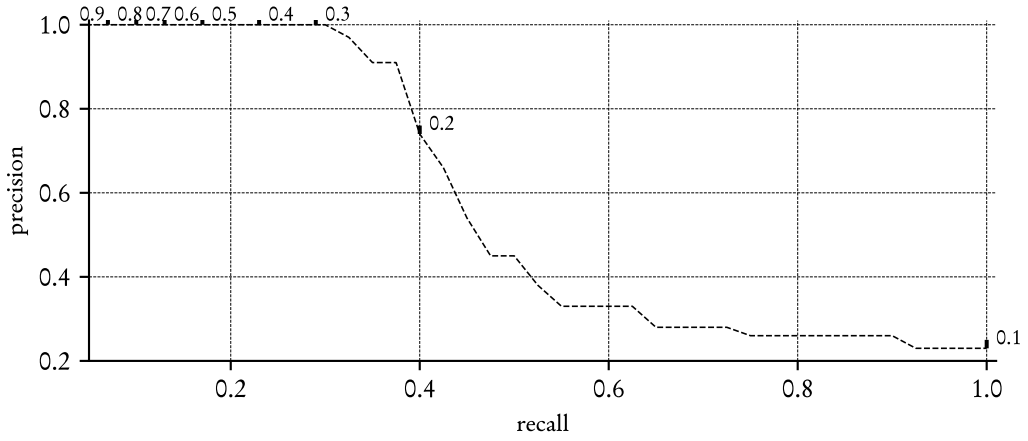



Figure 6.2: PR-curve for the *impostors* method with different $\sigma^*$ thresholds in a preliminary training round for 500-word samples of Hildegard of Bingen, Guibert of Gembloux and Theoderic of Echternach (tfidf-weighted vectors with standard-scaling). An ideal trade-off between recall and precision was found at a $\sigma^*$ value of 0.22. Evaluations can be found in table 6.1.

algorithm draws a decision boundary in between the candidates' works during training. Note that Theoderic and Guibert's works appear somewhat tricky to distinguish. Bringing in an additional, third component, as we will see in PCA plots further down this chapter, helped to make more nuanced distinctions between both authors, who apparently have quite a few stylistic aspects in common.

## 6.4 Results

The results of performing the rolling *impostors*-SVM method on the *Vita* are given in fig. 6.4. The x-axis shows the gradual progression of the sliding windows throughout the *Vita* by taking a step of 100 words at a time. The 'Corpus development' line at the very bottom of the figure indicates by chapter number in Klaes's edition which part of the *Vita* is treated per sample. Bars above the x-axis indicate the prediction of the *impostors* method, bars below the x-axis indicate the prediction of the majority of SVM classifiers. The y-axis (the height of the bars) indicates the confidence score (between 0 and 1) for both the *impostors* method and the SVM classifiers (longer bars with higher colour intensity indicate more confident attributions).

### 6.4.1 Hildegard's Autobiographical Fragments

The indexes 1–15 indicated below the figure are referenced in table 6.2. They mainly correspond (with a few exceptions discussed below) to Hildegard's autobiographical fragments, of special interest to us here. Immediately, it appears that despite the short sample length the combined method (*impostors*-SVM) recognizes the eight visions and other passages in which Theoderic cites Hildegard as strongly Hildegardian. The *im-*

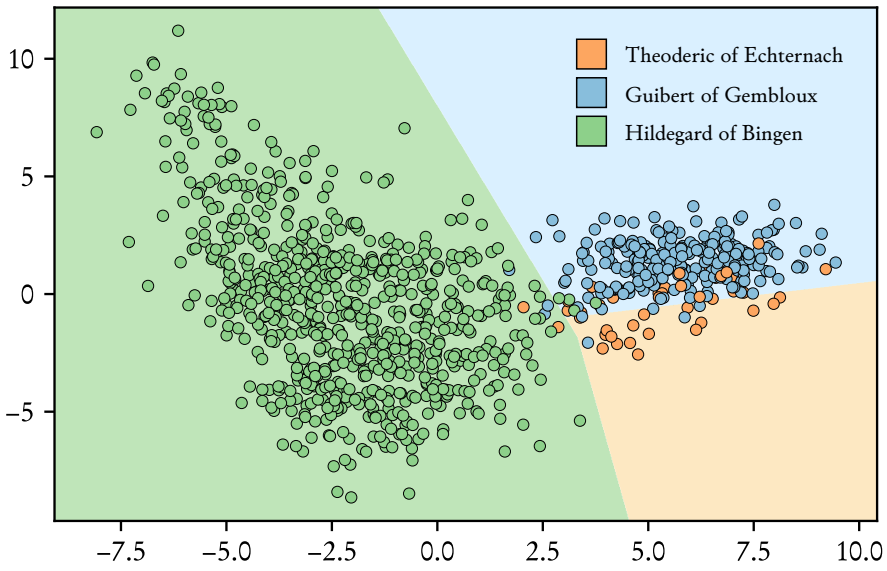



Figure 6.3: PCA plot giving a 2D intuition of the decision boundaries drawn between Theoderic, Guibert and Hildegard by the best-scoring SVM classifier. Settings: *s-l* = 500 w | *type* = most-frequent words | *n* = 500 | *vect.* = tfidf-weighted raw frequencies | *expl. var.* = 5.20%.

*postors* method is —as was to be expected— somewhat more severe in its prediction. When thrown in an 'open setting' (*impostors* method) the autobiographical samples are struggling far more to beat the competition by authors from the benchmark corpus. In a 'closed setting' (SVM classifier) they univocally adhere to Hildegard's style. 78 out of 140 samples assign Hildegard as candidate for the *Vita*'s authorship. This is an extensive and convincing contribution to the whole, and may counter skepsis on the authenticity of these passages, here meaning: their similarity to Hildegard's canonical works, which we presume to be the most reliable specimen of what constitutes Hildegard's style. In other words, the *Vita* may rightfully be designated 'autobiographical,' and Guibert's and Theoderic's influences —to which we will return below— remain limited.

A fascinating instance of an extremely small (!) portion of the *Vita* similar to Hildegard's style appears at the close of book 3's series of *miracula*. It concerns two letters, indexed as 14 and 15 in fig. 6.4, written by Hildegard's sisters:

> Now that we have edited everything as far as the capacity of our limited talent allows, let us turn our pen to the words of her holy daughters, who have written worthily of her memory. With the help of the Lord let us append to this work faithfully and truthfully what they saw and heard, especially concerning her blessed passage from this life, which they have written down with their own hands.[66]

[66] Translation taken from Godefridus Sancti Disibodi and Theodericus Epternacensis, *The Life of Hildegard*, 208; original text: "His - prout possibilitas ingenioli suppetebat - a nobis digestis calamum ad uerba sanctarum filiarum eius uertamus, et que de ipsa memoratu digna scripserunt, maxime de beato transitu eius, sicut uiderunt et audierunt et manibus suis tractauerunt, adiuuante Domino fideliter et ueraciter huic operi annectamus," see *Vita Hildegardis*, 3.26, ll. 10–4, 68.

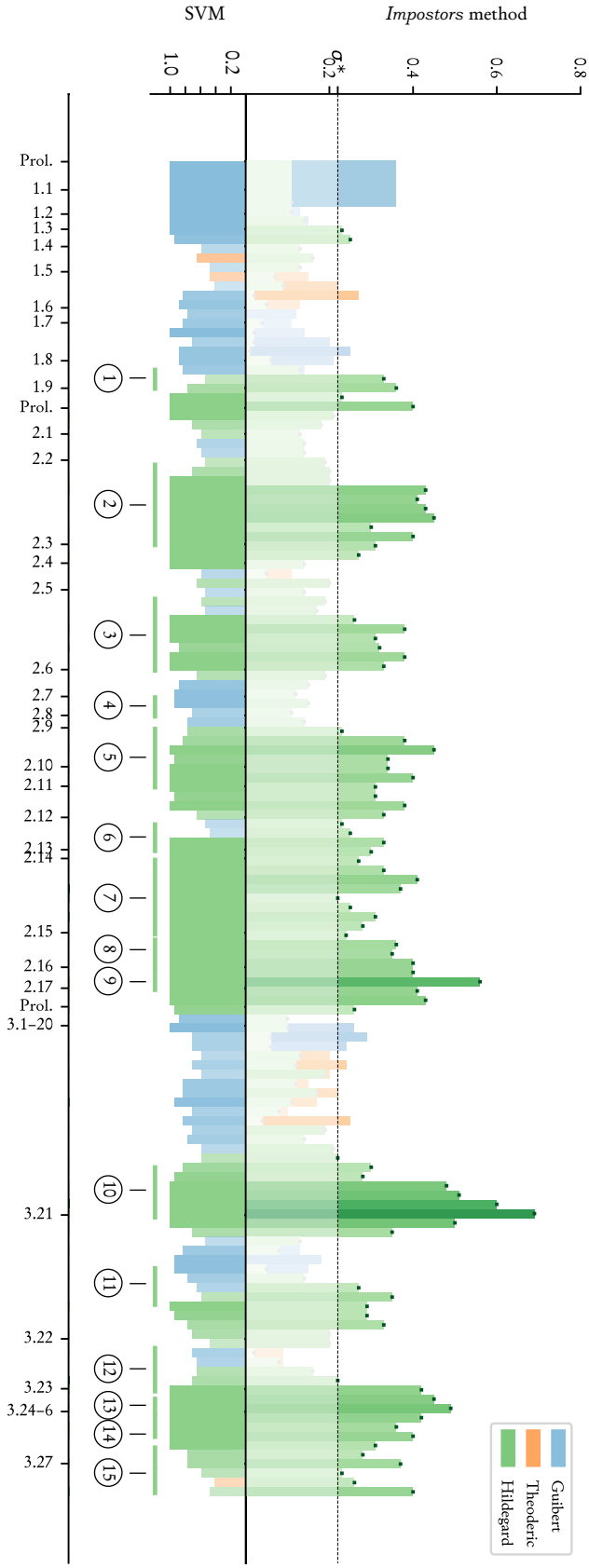



Figure 6.4: Rolling *impostors*-SVM method for the *Vita Hildegardis*. The x-axis shows the progression of the sliding windows throughout the *Vita* by taking a step of 100 words at a time. The 'Corpus development' line at the very bottom of the figure indicates by the chapter numbers in Klaes's edition which part of the *Vita* is treated. In total, attributions were made for 140 partially overlapping text samples consisting of 500 words. Bars above the x-axis indicate the prediction of the *impostors* method, bars below the x-axis indicate the prediction of the majority of SVM classifiers. The y-axis (the height of the bars) indicates the confidence score (between 0 and 1) for both the *impostors* method and the SVM classifiers (longer bars with higher colour intensity indicate more confident attributions). The indexes 1–15 given below the figure are referenced in table 6.2.

| Index | Fragment | Incipit | Ed. (M. Klaes) |
|---|---|---|---|
| 1 | Letter to Guibert | "Deus inquit ubi ..." | §1.8–9, pp. 14–5. |
| 2 | *Prima visio* | "In mystica inquit ..." | §2.2, pp. 21–4. |
| 3 | *Secunda visio* | "Quodam inquit tempore ..." | §2.5, pp. 27–30. |
| 4 | *Visio tertia* | "Vidi in visione ..." | §2.7, pp. 31–2. |
| 5 | *Visio quarta* | "In lectum egritudinis ..." | §2.9, pp. 33–5. |
| 6 | The philosopher | "Quidam phylosofus de ..." | §2.12, pp. 37–8. |
| 7 | *Visio quinta* | "In vera inquit ..." | §2.14, pp. 38–41. |
| 8 | *Visio sexta* | "Tres turres in ..." | §2.15, pp. 42–3. |
| 9 | *Visio septima* | "Subsequenti demum tempore ..." | §2.16, pp. 43–4. |
| 10 | Sigewize | "Posteaquam me visio ..." | §3.20, pp. 56–7. |
| 11 | Letter to Gedolphus | "G. ecclesie Brunwilarensis ..." | §3.21, pp. 60–2. |
| 12 | Possessed woman in Rupertsberg | "De adventu inquit ..." | §3.22, pp. 64–5. |
| 13 | Account of her illness | "Post hec inquit ..." | §3.23, p. 66. |
| 14 | *Visio octava* | "Pulcherrimus inquit et ..." | §3.24, pp. 67–8. |
| 15 | *Mulierem inquiunt* and *Cum beata* | "Mulierem inquiunt" \| "Cum beata" | §3.26–7, pp. 68–70. |

Table 6.2: Contents of fragments of the *Vita Hildegardis* indexed in fig. 6.4, completed with a description of their contents, incipit, and reference to the edition.

The first account of the sisters is an anecdotal and concise summary of miracular deeds performed by Hildegard (*Mulierem inquiunt*). The second is referenced in table 6.2 as *Cum beata*, and is preoccupied with Hildegard's illness and her death at the age of eighty-two, which is portended by the apparition of a glowing red cross at the firmament. After these two passages, at the very ending of the *Vita*, an additional, short unintroduced passage on Hildegard's burial occurs, for which Theoderic mentions no source in the text. It recounts the miracular benefits that visitors had gathered from venerating Hildegard's grave.

In the course of these final passages, the rolling *impostors*-SVM algorithm signalizes a lot of 'Hildegardian' material in the language. Evidently this poses a problem, as the passages include a description of events not only before but also after the author's death. Also on closer inspection of *Mulierem inquiunt* and *Cum beata*, one gains a strong impression that Hildegardian language is present. In the majority of turns of phrase one finds her preferred syntactic constructions and imagery, especially of *Scivias* and the *Liber divinorum operum*. I have appended a more detailed study of corresponding passages in A.6.1 and A.6.2 (p. 363), which were automatically searched by using Levenshtein distance.[67] A large number of sentences has parallels with passages in Hildegard's writings, especially the description of Hildegard's death in *Cum beata*, in which the red cross illuminates the sky into a colourful and dizzying spectacle.

We might want to pause briefly at what is happening here. One should not forget that Hildegard's *Vita* is classified as an autohagiography, a genre heavily based upon literary precedents. This is a text largely disinterested with factual reality or historical

[67] Levenshtein distance is a very simple operation for measuring the difference between two string sequences. A low Levenshtein distance means a close match between two word groups or sentences.

truth.[68] Hildegard's death, —as would the events leading up to it— would have to a large extent been pre-written according to the rules of the genre. The depiction of a saint's death was a literary topos, invoked with a specific purpose: the ultimate authentication of the saint's holiness. The conventional nature of death passages in female saints' hagiographical literature is an important point emphasized by Garay and Jeay in their recent piece exploring the "stages and staging of holy women's death." By discussing the death passages of female mystics such as Elisabeth of Schönau, Douceline of Digne († 1274), Marie of Oignies († 1213) and Lutgard of Aywiéres († 1246), they stress that "death is the moment which epitomizes the heroic life of women who have been chosen for the vocation of sainthood."[69] Many of the aspects Garay and Jeay attend to in order to expose the constructed nature of these death passages may well be shown to apply to Hildegard's *Mulierem inquiunt* and *Cum beata* as well. One is, for instance, the divine endorsement crucial to legitimating the saintly status: "But God showed clearly in her passing what standing she had before him."[70] Another is Hildegard's performance of "posthumous appearances and miracles":[71] "two men [...] made bold to touch her holy body" and "recovered from a severe illness."[72] Thirdly, the ending of the *Vita* allocates a large role to the participation of Hildegard's community. The posthumous miracles lead up to her enshrinement "in a venerable place," which draws pilgrims for its "many benefits [...] available to all who come seeking them with devout heart."[73] All of these elements make the depiction of Hildegard's death symbolically coincide with a wider involvement of the members of her community, for whom the cultivation of her person and the tradition she had founded becomes paramount. The death passage was, in other words, the apogee of the narrative, with a lasting importance for Hildegard's remembrance and for the economical survival of the Rupertsberg. It was, in Dalarun's words, "the transition of her personal charisma to a durable institution."[74]

[68] Kate Greenspan, "Autohagiography and Medieval Women's Spiritual Autobiography," in *Gender and Text in the Later Middle Ages*, ed. Jane Chance (Gainesville: University Press of Florida, 1996), 216–36.

[69] Kathleen Garay and Madeleine Jeay, "Sanctification of the Body: The Stages and Staging of Holy Women's Death in High Medieval Europe," in *Heroes and Saints: The Moment of Death in Cross-Cultural Perspectives*, ed. Phyllis Granoff and Koichi Shinohara (Newcastle-upon-Tyne: Cambridge Scholars, 2007), 139.

[70] "Deus uero, cuius meriti apud se esset in transitu suo euidenter declarauit," in *Vita Hildegardis*, 3.27, ll. 16–7, 70. Translation taken from Godefridus Sancti Disibodi and Theodericus Epternacensis, *The Life of Hildegard*, 209.

[71] Garay and Jeay, "Sanctification of the Body," 139.

[72] "Nam duo homines, qui sanctum corpus eius spe bona tangere presumpserunt, a graui infirmitate conualuerunt," in *Vita Hildegardis*, 3.27, ll. 37–9, 70. Translation from Godefridus Sancti Disibodi and Theodericus Epternacensis, *The Life of Hildegard*, 210.

[73] "Exequiis igitur uenerabiliter a reuerendis uiris celebratis in uenerando loco est sepulta, ubi meritis eius omnibus pio corde querentibus prestantur beneficia multa," in *Vita Hildegardis*, 3.27, ll. 39–41, 70–1. Translation from ibid.

[74] "Ce qui se joue en effet au moment précis du *transitus* du saint fondateur, c'est non seulement le passage attendu de l'ici-bas à l'au-delà, c'est aussi le passage périlleux d'un charisme personnel à une institution faite pour durer, d'un idéal toujours et oujours plus idéalisé par l'hagiographie à une pratique quotidienne, à une nécessaire insertion dans l'Église et dans la société," see Jacques Dalarun, "La mort des saints fondateurs. De Martin à François," in *Les fonctions des saints dans le monde occidental (IIIe-XIIIe siècle) Actes du colloque de Rome (27-29 octobre 1988)* (Rome: École Française de Rome, 1991), 194. Dalarun was speaking specifically of reformers, founders of orders and monasteries. Much of the same

Taking into account the foregoing, there is evidently more than one hypothesis which could account for why Hildegardian language appears here. The boldest one is to believe that Hildegard described the miracles in *Mulierem inquiunt* herself, and prophesied on the events of her death in *Cum beata* and ultimately arranged for the texts to be incorporated in her autohagiographical *Vita*, all of which occurred under her own authority and by her own hand. One may invoke one or two reasons in this hypothesis's defence. The *Vita* portrays Hildegard as prescient of the conditions by which she was to die, and, most importantly, as portending this course of events to her fellow sisters —to whom Theoderic emphatically attributes the authorship of the current passages.[75] The 'fabricating' of death stories has precedents in the twelfth century. One could think, for instance, of Geoffrey of Auxerre's death letter of Bernard of Clairvaux, composed in order to recuperate the saint's authority and authorize Arnaud of Bonneval's contribution to Bernard's *Vita*.[76]

However, it is a curious theory to believe that Hildegard deliberately sat down to write about her own death (amongst other matters), with her secretaries as accomplices to what can arguably be called a very bizarre undertaking indeed. A more acceptable hypothesis is that the method very subtly picks up on the fact that Hildegard's fellow sisters were trained extensively to imitate their *magistra*, and did an express effort to conjure up her style and tone in a passage with such great symbolic significance. Hildegard's words reverberate almost literally (again, I refer the reader to the table of correspondences on pp. 363ff.). Another hypothesis could be that the passage as a revision of authentic Hildegardian materials, recycled to an extent sufficient enough to fool the *impostors* method. By principles similar to those of end redactor Theoderic of Echternach, her sisters loosely collected some of Hildegard's remaining writings after her death in the Rupertsberg scriptorium and cobbled them together. Considering how the *Vita*'s composition process was one of recuperating materials that coincidentally happened to be at disposal, this may well be feasible.

could be argued for Hildegard, who founded the Rupertsberg, and definitely stood at the beginning of a female-oriented, visionary tradition.

[75] The saint's prescience on his or her death is, however, an often encountered hagiographical trope. See Pierre Boglioni, "La mort dans les premières hagiographies latines," in *Le sentiment de la mort au moyen âge*, ed. Claude Sutto (Montreal: L'aurore, 1979), 189; the original passage in Latin says "spiritu prophetie ei reuelauit, quem et sororibus predixit." See the *Vita Hildegardis*, 3.27, ll. 7–8, 69. Theoderic's abridged version states "Hec obitum suum longe ante presciens et sororibus predicens" in Theodericus Epternacensis, *Octo lectiones in festo sancte Hildegardis legende*, in *Vita sanctae Hildegardis virginis*, ed. Monika Klaes, *CC CM* 126 (Turnhout: Brepols, 1993), §8, ll. 15–6; and Guibert's revised version —with his words indicated between angle brackets— says "Cum beata inquiunt mater <regi et dominatori omnium> multis laborum <et dolorum> certaminibus deuote militasset, presentis uite tedio affecta <ad gaudia summe beatitudinis anhelans> dissolui et esse cum Christo cottidie cupiebat. <Quapropter> Deus <hanc dilectam suam a bono desiderio suo fraudari nolens diutius> finem <mortalis uite, quem ad ipsum suspirando optauerat,> spiritu prophetie ei reuelauit, quem et <filiabus suis in breui futurum esse sepe dixit>," see Guibertus Gemblacensis, *Vita sanctae Hildegardis retractata*, in *Vita sanctae Hildegardis*, ed. Monika Klaes, *CC CM* 126 (Turnhout: Brepols, 1993), 3.26, ll. 1–8, 106.

[76] Adriaan Hendrik Bredero, "Der Brief des heiligen Bernhard auf dem Sterbebett: eine authentische Fälschung," in *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, vol. 5, Monumenta Germaniae Historica Schriften 33 (Hannover: Hahnsche Buchhandlung, 1988), 201–24.

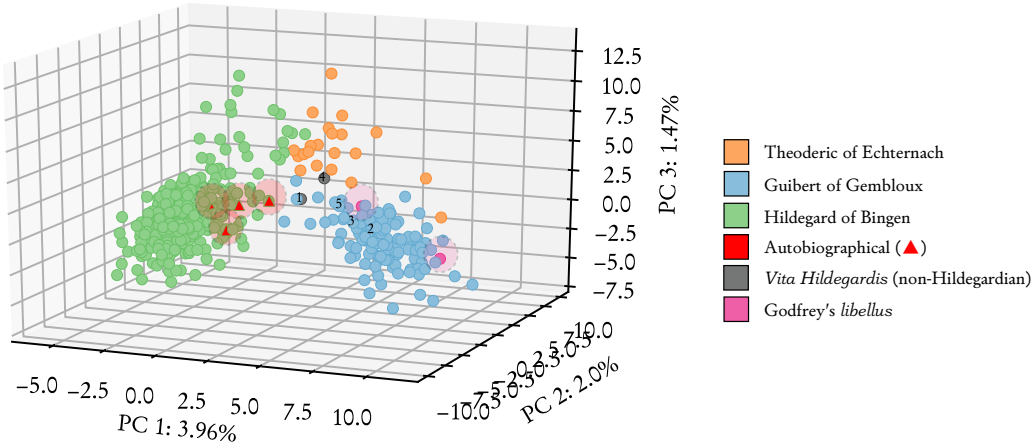



Figure 6.5: PCA plot containing Hildegard's, Theoderic's and Guibert's training texts (green, orange, blue). Theoderic's commentaries in the *Vita* are coloured gray, and come annotated with sample indices indicating a rough order of appearance (1–5). Also included in separate colouring are Godfrey of Disibodenberg's *libellus* or book 1 of the *Vita* (purple), and the autobiographical passages of Hildegard (red triangles), giving further evidence in support of fig. 6.4 that Hildegard's visions are fully Hildegardian. Settings: *s-l* = 1,000 w | *type* = most-frequent function words | *n* = 500 | *vect.* = tfidf-weighted raw frequencies | *expl. var.* = 7.43%.

### 6.4.2 Theoderic and Guibert

In the spirit of gradually moving from one inference to the next, we may make the assumption —based on the rolling *impostors*-SVM method above— that the autobiographical passages are genuinely Hildegard's. This is further confirmed in the PCA plot in fig. 6.5, where the original *Vita Hildegardis* was divided into two distinct batches: Hildegard's visions vs. all non-Hildegardian fragments of the text.[77] The division into batches also allowed this additional verification to work with 1,000-word instead of 500-word samples. Again, Hildegard's autobiographical passages can patently be shown to be Hildegard's (red triangles), and are clearly distinguishable from remaining samples of the *Vita*, namely Theoderic's commentaries (dark gray) and Godfrey's *libellus* (purple).

The behaviour of these remaining samples of the *Vita*, traditionally believed to have been the work of Godfrey of Disibodenberg and Theoderic of Echternach, prove far more difficult to categorize. If we revisit the predictions of the rolling *impostors*-SVM method earlier (fig. 6.4), and combine them with the PCA plot in fig. 6.5, the following indications are given:

1. The PCA plot in fig. 6.5 has Theoderic's commentaries (gray) and Godfrey's *libellus* (purple) cluster predominantly on the right end of the figure, alongside

[77] Note that the distinction was made on the basis of textual structure (i.e. we collected the text that fell under such headings as *Prima visio*, etc.) and not in strict accordance with the rolling impostors-SVM method's predictions (e.g. all samples predicted to be Hildegard's and coloured green in fig. 6.4.

the works of Guibert of Gembloux (coloured blue).

2. In a closed setting (SVM), (fig. 6.4) these samples —be it hesitantly— sympathize with Guibert as well.

3. The *impostors* method (fig. 6.4), on the other hand, refuses to become very confident, and makes few to almost no attributions to either Theoderic or Guibert which surpasses the $\sigma^*$ threshold.

That Theoderic's commentaries fail to cluster with any of his chronicles is particular, and calls for some additional analysis. The three PCA plots given in fig. 6.6 leave out Hildegard's works, and benchmark test documents Godfrey's *libellus* and Theoderic's commentaries against exclusively Theoderic's and Guibert's training texts (both individually and together). Here again, one gains the impression that the remaining samples of the *Vita* are inbetweeners, with a more outspoken preference to side with Guibert, be it never quite convincing. Klaes's indications that chapters §1.8–9 of the *Vita* (moreover containing the letter to Guibert), testify more of Theoderic's style, is not confirmed.[78]

None of both candidate authors are very convincing, and it turns out that Guibert is systematically the best guess, if guessing is at all warranted in this scenario. Even Theoderic's first prologue, in which he explicitly announces his presence and informs his readers on which source materials were used (without mentioning Guibert), turns out to be more like Guibert than like Theoderic. Guibert's (quite extensive?) stylistic influence on the *Vita* as we have it is problematic, for it does not appear compatible with the commonly accepted timeline of the *Vita*'s composition.

The prologues are important in establishing the chronology, for they give firm evidence of the current *Vita*'s completion by Theoderic, at a time definitely after Hildegard's death († 1179). It has commonly been assumed that Theoderic and Guibert just missed each other at the Rupertsberg. Theoderic arrived in the early 1180s shortly after Guibert had departed for Gembloux. Consequently the two biographers are thought to have been unaware of each others' *Vitae* until Guibert coincidentally discovered Theoderic's in corresponding with Godfrey of Saint-Eucharius (all of which explained earlier on pp. 182ff.), after which he revised it c. 1208/9 before including it in his own *epistolarium*. In their current form it must stand beyond doubt that the *Vita*'s prologues' *terminus a quo* is 1181, when Godfrey became abbot of Saint-Eucharius, and their *terminus ante quem* before Ludwig of Saint-Eucharius passed away, that is 1187.[79]

In other words, that Guibert's presence is suggested even in those passages of the text which have always been thought to have been exclusively Theoderic's additions is

[78] These chapters are collected under sample 2 (purple). For Klaes's suggestions of Theoderic's authorship, see *Vita Hildegardis*, 92*–94*.

[79] The prologues testify of Godfrey and Ludwig's simultaneous abbacy. See Silvas, *Jutta & Hildegard*, 121.

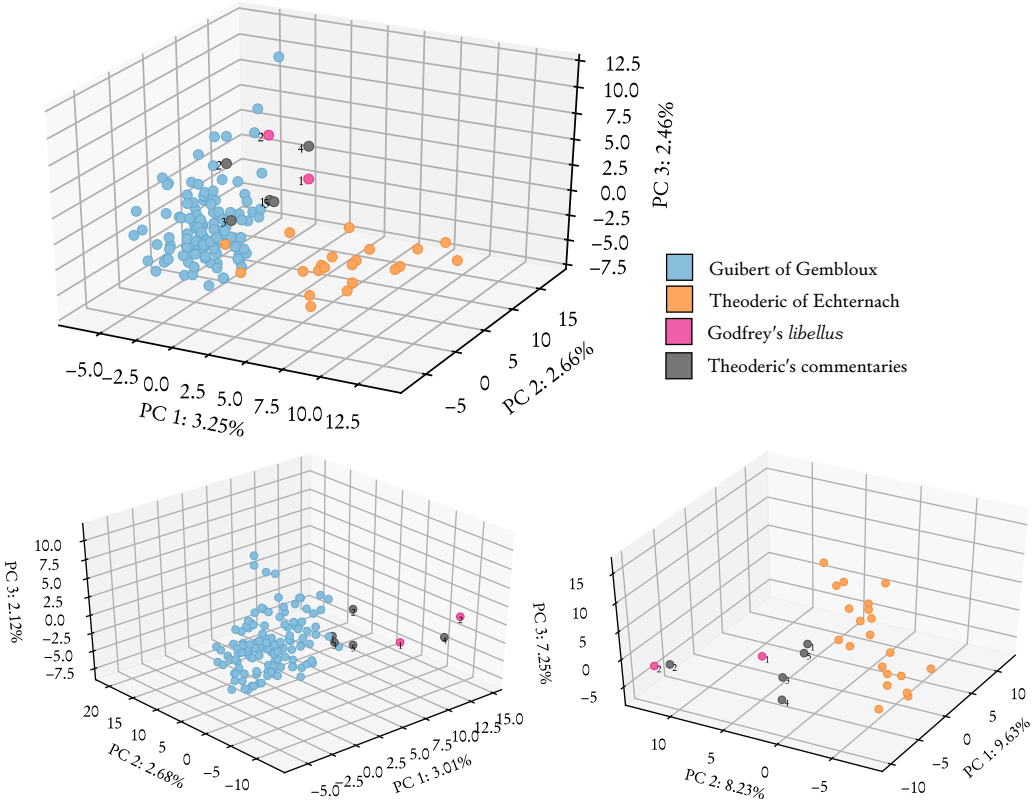



Figure 6.6: PCA plots of Theoderic's commentaries (gray) and Godfrey of Disibodenberg's *libellus* or book 1 of the *Vita* (purple). Theoderic's commentaries and the *libellus* are annotated with sample indices (1–5 and 1–2), indicating a rough order of appearance. Settings: *s-l* = 1,000 w | *type* = most-frequent function words | *n* = 500 | *vect.* = tfidf-weighted raw frequencies.

troublesome. I see two (maybe three) plausible hypotheses for explaining it, but hard historical evidence for either of them is lacking. Either the last redactor of the *Vita* was Guibert instead of Theoderic, or else Theoderic's reliance upon Guibert's source materials is far more extensive than has hitherto been presumed. Before discussing the pros and cons of either of them somewhat more extensively, it should be noted that both hypotheses are weakened by Guibert's seemingly genuine surprise upon learning in 1208/9 from abbot Godfrey that there is an extant *Vita* of Hildegard. If the *Vita* sent to him by Godfrey had been a work largely reliant upon his own text, then Guibert shows no sign of indignation or familiarity, no reaction at all really. Instead, Guibert responds to Godfrey's request for corrections by remarking that "I have nothing in memory to infer or add, nor can I find anything superfluous which I would remove, nor anything ineptly placed that I would correct."[80] It is peculiar to believe Guibert is talking about a *Vita* he had a hand in himself, unless we assume that these lines are intended to be tongue-in-cheek or slightly smug, or that he no longer

[80] "[...] Non habens pre memoria quid inferrem uel adderem, nec inueniens in ea quicquam superfluum quod demerem, neque aliquid inepte positum quod corrigerem [...]," see Guibertus Gemblacensis, *Epistolae*, Ep. 42, ll. 117–9, 394.

recognized his work after some thirty years, or that there were reasons for him to conceal his former contribution to this older *Vita*.

Let us explore the first hypothesis' presumptions and its merits. It presupposes that some time after Theoderic's completion, Guibert has seen a chance to extensively revise an earlier version of the *Vita* by the former's hand, which is now lost. This poses numerous problems. Firstly, the manuscript on which Klaes's edition is based has firmly been retraced to Echternach, and has on palaeographical grounds been shown to contain handwriting similar to that of Theoderic. In other words, our best manuscript of the *Vita* is an autograph by Theoderic.[81] One could always assume that Guibert visited the Rupertsberg while Theoderic was working on the *Vita* in the early 1180s, but this seems unlikely considering Guibert's recent departure and busy schedule, including a pilgrimage to Tours.[82] In his letter to abbess Ida, written around 1185, Guibert seems to be apologizing for his longstanding silence toward the Rupertsberg community after his departure, which he defends by stating that he had been victim of false accusations and jealousy toward him after his stay there.[83] Considering Guibert's close involvement in collecting the source materials, one might wonder why neither Ida nor any other Rupertsberg sister felt it necessary to inform Guibert on a new *Vita* in the works, or sent it to him if it happened to be finished around that time. Aside from these problems, the hypothesis that the *Vita* as we have it has known revisions by Guibert that postdate Theoderic's version becomes difficult in light of the fact that we already have a revision by Guibert, the *Vita sanctae Hildegardis retractata*.[84] This would lead to the conclusion that Theoderic's *Vita* contains Guibert's first revision, and that Guibert's *Vita retractata* is the revision of the revision.

The second hypothesis holds that we have underestimated the degree by which also Theoderic's interbeddings are heavily indebted to text prepared by Guibert between Hildegard's death and the latter's arrival at the Rupertsberg (1179/8). Guibert had stood in close contact with Hildegard from 1175 onward and had become her closest secretary in 1177. Being closely involved in the composition of her *epistolarium* and the completion of the Riesencodex, the *Vita* which Theoderic found upon arrival might have looked very similar to the one lying before us today. After all, who else would have found it more necessary to extend Godfrey's *libellus* with Hildegard's letter to Guibert, the *De modo visionis sue*, than Guibert himself (indexed as 1 in fig. 6.4)?[85] Also in his fragmentary *Vita* sent to Bovo (Ep. 38 of his letter collec-

[81] MS Vienna, ÖNB, 624. "Die eigentliche Provenienz des Kodex ist aber das Kloster Echternach, wo er vom Autor der Vita S. Hildegardis, dem Echternacher Mönch Theoderich, selbst aufgezeichnet wurde." See Klaes, *Vita Hildegardis*, 158*.

[82] Moens, "De horizonten," 76.

[83] Guibertus Gemblacensis, *Epistolae*, Ep. 32, ll. 100–113, 336. For all the letters with the Rupertsberg, see Ibid., Epp. 32–7, 333–65.

[84] Guibertus Gemblacensis, *Vita sanctae Hildegardis retractata*.

[85] On the *Vita*'s incorporation of this fragmentary letter (partly taken from Ep. 103$^{R}$ in Hildegard's letter collection), see n. 42.

tion),[86] Guibert included this letter from Hildegard, which shows the importance he attached to it. Following this train of thought, we may assume that Theoderic made subtle stylistic amendments, perhaps inserted references to Echternach's well-known abbot Thiofrid,[87] but in reality heavily relied —including even large parts of the three prologues to the individual books— on an architecture formerly constructed by Guibert (and, perhaps, also Hildegard). Theoderic but had to score out Guibert's name, insert the necessary realia, and assemble the entire work under his name so as to finish the task. That Theoderic did not name Guibert as his predecessor is reminiscent of how Guibert had himself erased the existence of his predecessor Godfrey, so as to enhance his position as direct successor of Volmar.[88]

From what the sources tell us when it comes to Theoderic's final redaction, which was to assemble the pre-existent material, this second hypothesis wins my personal favour, although substantial weaknesses remain. The question rises why Guibert would have left Theoderic a *Vita* in such an advanced stage of completion, although 'completion', of course, is a relative term in the Middle Ages. To him, whatever work he left behind in the Rupertsberg had been unfinished.[89] Here again, it has always been assumed that whatever work Guibert had started on a *Vita* during his time at the Rupertsberg is contained within the fragment sent to Bovo.[90] Why are Theoderic's *Vita* and Guibert's fragmentary *Vita* so dissimilar, if we suspect that Guibert was at the origin of both of them? And finally: if Guibert's influence to the whole was as extensive as I am insinuating, then why does he appear —as the current experiments have shown— to have remained loyal to Hildegard's source material instead of extensively revising it?

These questions are bound to remain open for now. One might be excused to ask how far one is willing to go in speculation, if these results do not provide a better timeline than that of Klaes, or if they might simply be confronting us with the limits of what is methodologically feasible. Perhaps we are handling a collaborative style so far advanced that stylometry abandons us. The *impostors* method's suggestion is better taken seriously: there are simply no favourite candidates amongst all the authors included in the benchmark corpus. The many hands' involvement in a *Vita* undoubtedly important for many of Hildegard's close followers might defy the detection of single-author stylistic elements.

[86] Guibertus Gemblacensis, *Epistolae*, Ep. 38, 366–79.

[87] See Klaes, *Vita Hildegardis*, 84*.

[88] Schrader and Führkötter, *Echtheit*, 147-50.

[89] So much becomes clear from his letter to abbot Godfrey: "opus ceptum imperfectum reliquisse." See Guibertus Gemblacensis, *Epistolae*, Ep. 42, l. 144, 394.

[90] Guibertus Gemblacensis, *Epistolae*, Ep. 38, 366–79.

## 6.5 Conclusive Remarks

Whereas we have begun this chapter by questioning Hildegard's autobiographical fragments in the *Vita*, we have instead ended with new questions concerning Theoderic's and Guibert's respective contributions to the interbedding commentaries, where much remains unclear. It turns out that Hildegard's autobiographical fragments appear uncorrupted despite their transmittal through the hands of multiple biographers. This is the only result in this chapter that I believe can stand as conclusive. The experiments' remaining results, however, mainly give indications toward further investigation.

One of them is the observation that the two letters by Hildegard's sisters reporting on her death and appended to the *Vita*'s third and final book, *Mulierem inquiunt* and *Cum beata*, are heavily indebted to Hildegard's wording and imagery. Either the sisters of the Rupertsberg meant to resuscitate Hildegard's tone and authority at the very end of her *Vita*, and perhaps even drew on Hildegard's materials so as to literally invoke her style, or perhaps Hildegard may even have had a hand in them herself. Undoubtedly, Hildegard's style was imitated at the Rupertsberg, where multiple of her assistants had been in the front row in learning to imitate and conjure up the visionary's style. Then again, that Hildegard was somehow involved herself is not impossible per se. She is known to have participated in collecting and revising her *opera omnia* during the last years of her life, which had the aim of representing her image for posterity, and the saint's death is a crucial culmination point for a project largely invested in canonization, in which Hildegard was meant to be depicted in a larger-than-life, hagiographical fashion. For what it is worth, also the *Vita* itself reported on how Hildegard dictated the events surrounding her death to her sisters.

Paradoxically, whereas Hildegard's authority was not undermined in the autobiographical fragments, the largest tussle for stylistic dominance appears to have taken place in the commentaries guiding them. These have commonly been taken to have been written by end redactor Theoderic of Echternach. However, Guibert of Gembloux's style appears present in a few of them, which might lead to suspect that either Guibert had opportunity to revise the *Vita* at a time when Theoderic was (near to) completing it, or else that Theoderic largely relied on preparations carrying Guibert's mark. The first argument is hardly sustainable when based on Guibert's whereabouts during the time of Theoderic's ending of the *Vita*, but there is something to be said for the latter hypothesis. Then again, to my knowledge there is no additional evidence to support it aside from the statistical suggestions in this chapter, wherefore the question remains open and no conclusive answer is possible yet.

# 7

# Forging Ties: Suger and the Donation of Charlemagne

## 7.1 Forgery in Twelfth-Century Saint-Denis

In this next chapter our focus sweeps back to northern France, more particularly to the abbey of Saint-Denis under the abbacy of Suger (1080/1–1151) and his entourage. Although Suger is to be contextualized in a monastic environment, as were Bernard, Elisabeth and Hildegard, Suger's sphere of influence extends somewhat more into an administrative and diplomatic milieu. This is not to say that Suger has not occasionally shown himself a competent man of letters (e.g. his *De consecratione*), but his true influence was of a political rather than a philosophical-literary nature. His works, then, have often been deemed to betray a style that is indebted to the formulas and writing habits of a chancery which relied on collaborative efforts.[1] In regard of this context, the stylistic analyses in this chapter will also shift toward a new type of text which had as yet remained undiscussed in this thesis: the charter. As we come to discuss Suger's style and those of his secretaries, and as we come to discuss a number of suspect charters that can be brought in connection to him, a secondary research question to this chapter will arise: can elements of personal style be detected in the style of formulaic, administrative documents, and vice versa?

The primary research question of this chapter, however, mostly harks back to the tension between the physical scriptor of a text as opposed to author as (s)he is proclaimed within the text. Already throughout the former chapters, the ways in which computational stylistics may contribute to this debate by 'unmasking' the presupposed role-playing within a text has been demonstrated. More specifically, emphasis was placed on how authors strategically impersonate or construct literary identities. That these questions on fact and fiction arise even more fervently in the context of the

[1] Françoise Gasparri, "La politique de l'abbé Suger de Saint-Denis à travers ses chartes," *Cahiers de civilisation médiévale* 46, no. 183 (2003): 236.

multi-talented and versatile Suger, and the abbey of Saint-Denis, is no coincidence.

The abbey of Saint-Denis was notorious for the number of forgeries and falsifications it produced, and for its successful interweaving of legend and reality.[2] Therefore, it has throughout its history attracted renowned philologists, historiographers and connoisseurs who took it upon themselves to distinguish the authentic from the inauthentic. Lorenzo Valla (1405/7–1457), one of stylometry's forefathers,[3] was the first since Peter Abelard to call into question the conflated identity of Saint-Denis' patron saint Dionysius the Areopagite.[4] Furthermore, we have Saint-Denis' fondness for producing forgeries and fakes to thank for the existence of diplomatics. As was likewise expounded upon (see p. 85 above), Jean Mabillon (1632–1707), the forefather of diplomatics, defended the authenticity of Merovingian charters of the abbey of Saint-Denis after having been challenged by the Bollandist Daniel Papebroch (1628–1714). This in its own turn triggered his writing of *De re diplomatica* (1681), a classic for Latin antiquarianism and contemporary diplomatics.[5]

Even though Suger and his secretaries will play a central role in this chapter, the complex position of Saint-Denis in twelfth-century France and the peculiarities of the text that will be analyzed further down demands us to opt for a different departure point than in the previous chapters. Instead of departing from the problematization of the composition context, the current chapter will set out by problematizing the origins of one suspect text in particular, namely a donation charter of Saint-Denis written under the name of emperor Charlemagne (747/8–814).

## 7.2 The False Diploma: D Kar 286

In his history of the abbey of Saint-Denis, published in 1625, dom Jacques Doublet (1560–1648) printed an early ninth-century Carolingian charter (*Diplomata Karolinorum* 286)[6] in which emperor Charlemagne, in the presence of a general council of

[2] On the general topic of forgery in the Middle Ages, see especially Constable, "Forgery and Plagiarism."

[3] Valla was introduced in this thesis's second chapter. See p. 84. Especially Lorenzo Valla's treatise on the *Donatio Constantini*, the notorious decrete in which Constantine would have allegedly transferred secular authority over to Rome after having been healed of leprosy by Silvester I, is cited in these stylometric studies, because Valla pays specific attention to the Latin style of the forger, especially smaller, peculiar words and constructions that do not seem to belong there. See Eder, "A Bird's Eye View of Early Modern Latin," 61.

[4] On the conflated identity of Saint-Denis's saint patron Dionysius, discussed in more detail in n. 26, see David E. Luscombe, "Denis the Pseudo-Areopagite in the Middle Ages from Hilduin to Lorenzo Valla," in *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, vol. 1, Monumenta Germaniae Historica Schriften 33 (Hannover: Hahnsche Buchhandlung, 1988), 149–51.

[5] Mabillon, *De re diplomatica*.

[6] The charter was supposedly issued in the year 813, see Jacques Doublet, *Histoire de l'abbaye de S. Denys en France* (Paris: Chez Jean de Heuqueville, 1625), 725–7; introduced with the heading "Charte du Roy Sainct Charlemagne, par laquelle appert de sa grande dévotion envers son Apostre le glorieux Sainct Denys & ses compagnons, pour la reverence desquels il restitue à l'Abbaye & Convent de Sainct Denys une grande quantité de biens, situez en divers pays & contrees du Royaume de France, lesquels avoient estè usurpez sur icelle." The diploma has also been reprinted in the *MGH*, in the series of charters of Pippin the Short, Carloman and Charlemagne. There it has been labelled as D Kar 286, see Engelbert Mühlbacher, ed., *Die Urkunden Pippins, Karlmanns und Karls des Grossen*, vol. 1, *MGH* (Hannover: Hahnsche Buch-

leading clergy and nobility, granted the church abbey of Saint-Denis the privileged position of "head of all the kingdom's churches" (*caput omnium ecclesiarum regni nostri*). In the charter, the king of the Franks sets a precedent by recognizing Saint Denis (Dionysius the Areopagite)[7] as patron saint and protector, stating that his successors can only be legitimately crowned in Saint-Denis, and that only through Saint-Denis' apostolic authority archbishops and bishops can be ordained in the future. In conclusion he lays down his crown and royal insignia on the altar of the Areopagite, and consequently pays a tribute of four golden coins to augment the church building. He orders all bystanding householders to follow that example, and solemnly declares that from here on out such will be common practice. When even Charlemagne, the most powerful man of the kingdom, owes fealty to the abbey of Saint-Denis, everyone subordinated to that man must likewise offer loan to the church.

The so-called donation of Charlemagne to Saint-Denis, as summarized above, is without question a myth, and charter D Kar 286 a forgery, one of the many that were originally contained in the *chartrier* of Saint-Denis.[8] Yet, strangely enough, this must practically be the only point at which scholars have established agreement. Hardly anything is known for certain about this text. Not when it was written —possible dating ranges from the late eleventh century to the seventeenth—,[9] by whom it was written —abbot Suger of Saint-Denis, his successor Odo of Deuil (1110–1162) or some other secretary, or the seventeenth-century Benedictine monk Jacques Doublet? —,

handlung, 1906), 428.

[7] See n. 26

[8] What the archives of Saint-Denis originally looked like in the High Middle Ages is unknown, and the transmission history by which most charters were transmitted to us today is highly complex. What is certain is that the bulk of the charters that must have originally been contained within the Saint-Denis *chartrier* is found in Paris, most notably in the national archives (Arch. nat.) and the national library of France, the *Bibliothèque nationale de France* (BnF). The collection of Saint-Denis, including some rare Early Medieval charters, was severely dismembered, especially as the original charters became scattered across various archives and libraries from the Early Modern period and the French Revolution onward. In the nineteenth century, the body of charters was further subjected to mutilation, as editors reclassified them in series that appertained to modern research interests and the charters' contents. Ever since, there have been attempts to reconstruct the original corpus, but a full recovery of the original is now beyond possible. See Daniel Sonzogni, "Le chartrier de l'abbaye de Saint-Denis en France au haut Moyen Age. Essai de reconstitution," *Pecia* 3 (2003): 9–211; and Florence Clavaud, "The Digital Edition of the Medieval Charters of the Abbey of Saint-Denis: First Results and Prospects," *Digital Medievalist* 8 (2013), doi:`http://doi.org/10.16995/dm.48`; on forgery in Saint-Denis, also see Patrick J. Geary, *Phantoms of Remembrance: Memory and Oblivion at the End of the First Millenium* (Princeton: Princeton University Press, 1994), especially chap. 3, 107–13.

[9] From earliest to latest, the conjectured datings of the diploma's confection have been: eleventh century or early twelfth century, see Hermann Grauert, "Novitätenschau," *Historisches Jahrbuch* 12 (1891): 173; c. 1124, see Robert Barroux, "L'Abbé Suger et la vassalité du Vexin en 1124. La levée de l'oriflamme, la Chronique du *pseudo-Turpin* et la fausse donation de Charlemagne à Saint-Denis de 813," *Le Moyen âge: bulletin mensuel d'histoire et de philologie* 64 (1958): 1–26; c. 1127–9, see Manfred Groten, "Die Urkunde Karls des Großen für St.-Denis von 813 (C 286), eine Fälschung Abt Sugers?," *Historisches Jahrbuch* 108 (1988): 9; c. 1140–3, see Christopher Hohler, "A Note on *Jacobus*," *Journal of the Warburg and Courtauld Institutes* 35 (1972): 31–80; c. 1147–9, see Max Buchner, "Das gefälschte Karlsprivileg für St. Denis B M.[2] Nr. 482 und seine Entstehung," Historisches Jahrbuch, 42 (1922): 2.13; c. 1165, see Mühlbacher, *Urkunden*; and Co Van de Kieft, "Deux diplômes faux de Charlemagne pour Saint-Denis, du XIIe siècle," *Le Moyen âge* 64, no. 4 (1958): 401–36; and the seventeenth century, see Elizabeth A.R. Brown, "Saint-Denis and the Turpin Legend," in *The Codex Calixtinus and the Shrine of St. James*, ed. John Williams and Alison Stones, Jakobus-Studien 3 (Tübingen: Gunter Narr Verlag, 1992), 51–88.

by what purpose it was written —is this a laughable hoax, or genuine usurpation of ecclesiastical power by the abbey of Saint-Denis?—, and how it stands in connection to other texts that have an agenda of corroborating Saint-Denis's privileged relationship with the French court, not least the anonymous chronicle that circulates under the title *Pseudo-Turpin* (named after Charlemagne's war companion, bishop Turpin of Rheims, who was most definitely not the author).[10]

There has thusfar been no trace of an existent medieval original of the donation of Charlemagne, whose codicological or palaeographical evidence might have eliminated a number of leads on its origins. Unfortunately, we have to make do with a most unhelpful *textus unicus*: the aforementioned published transcription of the text by dom Jacques Doublet, who was himself monk at Saint-Denis in the seventeenth century, and possibly entertained motives for manufacturing the charter himself.[11] Due to this lack of a paper trail —or, better, maybe, 'parchment' trail—, text-internal criticism has been the final and only resort for many scholars who strove to contextualize the charter.

Throughout the current chapter, we will return to how the debate on the notions of medieval truth- and falsehoods in the case of the donation of Charlemagne ties in with sensitive, pertinent questions that apparently fascinated generations of medieval scholars up until today. At the heart of the matter is the question of how writers in the Middle Ages negotiated the boundaries between what is fiction and what is jurisdiction (juris-fiction), if authority had priority over authenticity, and generally what was considered 'truth' in a medieval context. To what extent would the legitimization of a good story entail far-reaching, tangible consequences in everyday life? Does the forging of a text —be it a legal or literary text— imply the forging of a reality? And finally, who was the author of this meticulously orchestrated confusion: Suger of Saint-Denis, or his close confidants (and collaborators) Odo of Deuil and William of Saint-Denis?

### 7.2.1 The Cult of Charlemagne: Saint-Denis and the Capetians

Most scholars agree that D Kar 286 needs to be understood as forming one of many links in a complex chain of legendary texts revolving around the figure of Charle-

[10] The *Pseudo-Turpin* presents a vast and complex research field of its own, not in the least because the text was extremely popular during the Middle Ages and has been transmitted in more than 200 manuscripts. The text will be discussed more fully on p. 212 below. It was edited twice in the 1930s, the first Cyril Meredith-Jones, ed., *Historia Karoli Magni et Rotholandi ou Chronique du Pseudo-Turpin, textes revus et publiés d'après 49 manuscrits* (Paris: Librairie E. Droz, 1936); and the second Hamilton Martin Smyser, ed., *The Pseudo-Turpin, edited from Bibliothèque Nationale, fonds latin, MS. 17656 with an Annotated Synopsis*, The Mediaeval Academy of America 30 (Cambridge, MA: The Mediaeval Academy of America, 1937), needless to say that the text's publication in two independent editions —contradicting many of each others' stances— failed to clarify the problems regarding the *Pseudo-Turpin*'s origins and purpose ; the latest edition of the text can now be found in Adalbert Hämel and André de Mandach, eds., *Der Pseudo-Turpin von Compostela* (München: Verlag der Bayerischen Akademie der Wissenschaften, 1965).

[11] This is a conclusion formulated by Brown, "Saint-Denis and the Turpin Legend," we will return to her suspicions later on.

magne. These tales, appearing in the form of epics, poems, *vitae* and hymns that revolved around the expeditions of the Frankish emperor, knew a great revival throughout Western Europe in the twelfth century, ultimately resulting in Charlemagne's canonization in 1165 and the composition of an additional *Vita Karoli Magni* (not to be confused with Einhard's ninth-century text) at Frederick Barbarossa's (1122–90) instigation in 1166.[12] They narrated, for instance, how Charlemagne travelled eastward to Constantinople and Jerusalem[13] and brought home Christ's nail and crown of thorns as relics, or conquered over the Saracens and freed and converted all of Spain "with word and sword."[14]

These miracular histories, the *gesta* of Charlemagne —chansons de geste—, came with such an appeal that both secular and clerical powers had every interest in becoming stakeholders to them. The remembrance of Charlemagne gave rise to the inauguration of annual festivities and/or cultic rites, and the creation of devotional buildings and objects. The churches of Aachen,[15] Orléans,[16] Rheims[17] and many others[18] all strove to partake in the lustre of Carolingian culture. The Capetians nourished the belief ever since Hugh Capet's accession in 987 that their reign was a continuation and not an interruption of the Carolingian line.[19] On the other side of the Rhine, the Holy Roman emperors, amongst whom Frederick Barbarossa, would equally be

[12] The text is also known as *De sanctitate meritorum et gloria miraculorum beati Karoli Magni*, and is edited in Gerhard Rauschen, *Die Legende Karls des Grossen im 11. und 12. Jahrhundert. Mit einem Anhang über Urkunden Karls des Grossen und Friechichs I. für Aachen von Hugo Loersch*., Publikationen der Gesellschaft für Rheinische Geschichtskunde 7 (Leipzig: Verlag von Duncker & Humblot, 1890), 17–93.

[13] In fact Charlemagne never travelled to Constantinople or Jerusalem. Benedict of Saint Andrew's on Mount Soracte in Southern Italy was the first to have initiated the myth, probably around 998, see Ronald N. Walpole, "The *Pèlerinage de Charlemagne*: Poem, Legend, and Problem," *Romance Philology* 8, no. 1 (1954): 180.

[14] As phrased in the forged Aachen charter, "ad fidem catholicam verbo convertit et gladio." See Hugo Loersch, "Das falsche Diplom Karls des Grossen und Friedrichs I. Privileg für Aachen vom 8. Januar 116," in *Die Legende Karls des Grossen im 11. und 12. Jahrhundert*, 154–5.

[15] Aachen (Aix-la-Chapelle) was of course the city where Charlemagne had founded his church. The Paltzkirche is what remains today. The relics which Charlemagne brought from Jerusalem and Constantinople would have stayed in Aachen for some time, before Charles the Bald (843–877) transferred them to Saint-Denis. In 1165, Frederick Barbarossa issued Charlemagne's canonization there.

[16] Orléans had started to compete with Rheims and Saint-Denis for privileged ties with the monarchy. This trend was especially incited after Philip I's burial in the abbey church of Fleury (in Saint-Benoît-sur-Loire, located nearby Orléans), which violated the habitual custom of French kings' interment in Saint-Denis. Philip I's successor, his son Louis VI (1081–1137), was crowned likewise in Orléans in August 1108, which in its own turn went against the king's customary coronation in Rheims. These events stimulated Hugh of Fleury to write and dedicate a history of the French kings to Louis VI, "that set out Fleury's historical role in promoting royal legitimacy," see James Naus, *Constructing Kingship: The Capetian Monarchs of France and the Early Crusades*, Manchester Medieval Studies (Manchester, UK: Manchester University Press, 2016), 62; in addition, links to the monarchy and common origins with Charlemagne could also be made by appealing to Orléans' school institutions, which had their roots in the person of bishop Theodulf of Orléans (c. 750/60–821). Theodulf, formerly also abbot of Fleury, had been a close confidant to Charlemagne and co-author of the *Libri Carolini*, see Léopold Delisle, *Les écoles d'Orléans au douzième et au treizième siècle*, L'annuaire-bulletin de la société de l'histoire de France 7 (Paris: Imprimerie générale de Ch. Lahure, 1869), 1; in conclusion, one could also refer to yet another forged donation charter, allegedly written out by Charlemagne, composed in Saint-Euverte (nearby Orléans) to authenticate the relics of their patron saint Evurtius, see Thomas Head, *Hagiography and the Cult of Saints: The Diocese of Orléans, 800–1200* (Cambridge, UK: Cambridge University Press, 1990), 51–2.

[17] Rheims had ever since 816 been regarded the coronation church of French monarchs by virtue of possessing Clovis's

occupied by maintaining a connection to the Carolingian past of East Francia.[20]

The abbey of Saint-Denis is presumed to have had a most central position in the twelfth-century cultivation of the Carolingian past.[21] In this regard, it first and foremost needs to be emphasized that the motivations for Charlemagne's idolization should not merely be sought in the Sandionysian monks' folkloric sentiment. There were juridical, political and economical benefits to authorizing oneself through the legendary emperor.[22] To this end, parties such as Saint-Denis that strove to affiliate themselves with the emperor's name and fame did not flinch from manipulating and confecting objects and texts in such a way that it advantaged their present interests.[23]

In fact, already as early as the late eleventh century, Saint-Denis prided itself of possessing the relics of the Passion (the nails and the crown) which Charlemagne had brought home from his travels eastward. Charles the Bald, Charlemagne's grandson, would have ensured the relics' delivery to the abbey of Saint-Denis, an event that was commemorated annually in June through the exhibition of the reliquaries during the fair of Lendit.[24] Such authentications through a direct historical connection to Charlemagne were employed to strengthen Saint-Denis' present connection with the royal court as Charlemagne's successors, and as lead participants in Charlemagne's tales

(c. 466–511) holy ampulla. It was Charlemagne's son, Louis I (the Pious, 778–840), who had installed the custom. A striking instance of Rheims' emulation with Saint-Denis is its renovation of the cathedral at instigation of archbishop Samson of Mauvoisin († 1161), who completed the project in close imitation of Suger's renovations in Saint-Denis, see Meredith Parsons Lillich, *The Gothic Stained Glass of Reims Cathedral* (Pennsylvania: The Pennsylvania State University Press, 2011), 238. Also Charlemagne's associate Turpin (or Tilpin, † 794/800) could be mentioned. Turpin was archbishop of Rheims and was believed in the twelfth century —by papal authority— to have authored the *Pseudo-Turpin* containing Charlemagne's *Vita*.

[18] "By the High Middle Ages claims to Carolingian origins had become widespread throughout Europe," see Geary, *Phantoms of Remembrance*, 135.

[19] Gabrielle M. Spiegel, "The *Reditus Regni ad Stirpem Karoli Magni*: A New Look," *French Historical Studies* 7, no. 2 (1971): 145–74.

[20] The topic of the reception of Charlemagne in the medieval German sphere is obviously a topic too extensive to discuss here in full, therefore see Robert Folz, *Le souvenir et la légende de Charlemagne dans l'Empire germanique médiéval*, Publications de l'Université de Dijon 7 (Paris: Les Belles Lettres, 1950).

[21] Even though Waldman pointed out that the Sandionysian monks had no special relationship with the last Carolingians, and received no notable royal attention and favour after the early tenth century and before the beginning of the twelfth, see Thomas G. Waldman, "Saint-Denis et les premiers Capétiens," in *Religion et culture autour de l'an Mil*, ed. Dominique Iogna-Prat and Jean-Charles Picard (Paris: Picard, 1990), 191–7; also see Geary, *Phantoms of Remembrance*.

[22] Emperor Charlemagne had come to represent a lawgiving authority: "Das allgemeine europaweite Ansehen Karls des Großen, sein Ruf als Rechtsschöpfer und Rechtswahrer, hatten zur Folge, daß vielfach —ohne daß aus diesem Umstand zunächst oder gar immer eine materielle Verfalschung entstanden wäre— spätere Rechtssätze und Gewohnheiten mit dem Frankenkaiser in Verbindung gebracht wurden, eben aus der notorischen Uberzeugung des guten und alten 'Karls-Rechtes,'" see Dieter Hägermann, "Die Urkundenfälschungen auf Karl den Großen: Eine Übersicht," in *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, vol. 3, Monumenta Germaniae Historica Schriften 33 (Hannover: Hahnsche Buchhandlung, 1988), 440.

[23] The number of falsified charters written in Charlemagne's name is impressive, especially (yet not exclusively) in the eleventh and twelfth centuries. For a survey and discussion of the falsified Charlemagne charters, see ibid.

[24] It was a commonly held belief in the Middle Ages that the fair of Lendit was a centuries-old tradition initiated in the ninth century by Charles the Bald. However, the fair's origins are eleventh century (1048), and its annual rehearsal was initiated in the early twelfth-century, the first of which took place in 1109, see Léon Levillain, "Essai sur les origines du Lendit," *Revue Historique* 155, no. 2 (1927): 241–76.

that increasingly became widespread.[25] It consolidated Saint-Denis' supremacy over some of the churches of France that rivalled to entertain such a privileged relationship with the Capetians, not in the least Rheims. Saint-Denis' possession of the relics was narrated in the *Descriptio qualiter*, a late eleventh-century propagandic document describing the eastward travels of Charlemagne, by some believed to have originated in Saint-Denis. The text is a "description [*descriptio*] of how [*qualiter*] Charlemagne brought the nail and crown of thorns of the Lord from Constantinople to Aachen and how Charles the Bald then transferred them to Saint Denis."

This power relationship between Saint-Denis and the Capetians was one under constant negotiation, and the balance was prone to tip over. Saint-Denis saw its patron saint Dionysius the Areopagite[26] as liege and protector over the French court.[27] A critical juncture for Saint-Denis in stipulating the spiritual sovereignty of their patron, was king Louis VI's taking up of the red banner of Vexin from the high altar of Saint-Denis in 1124 when marching out against the Holy Roman emperor Henry V. Under the banner, king Louis VI (1081–1137) mobilized and united all the French nobility that mattered at the time. The banner, on which the patron Dionysius would have been depicted, symbolized the king's vassality towards the abbey of Saint-Denis.[28] For a long time, the banner of Vexin was identified as the *Oriflamme*, which Charlemagne carried to battle in the *Chanson de Roland*. Abbot Suger, who describes the event of Louis VI's war declaration in the *Vita Ludovici regis VI*, does not make such a direct identification, but the confusion of the two banners would have been customary by the end of the twelfth century:[29]

> Then [Louis VI] hurried to the blessed Dionysius, for common report and frequent experience had taught him that he was the particular patron and, after God, the foremost protector of the realm. [...] He then took from the altar the standard belonging to the county of Vexin, which he held as a fief from the church, and gazed upon it. Taking it

[25] The text is likewise edited in Rauschen, *Die Legende Karls des Grossen*, 103–25.

[26] Saint-Denis' patron, Dionysius the Areopagite, was in medieval times a conflation of three figures to one, namely Dionysius the Areopagite, the first-century converted bishop of Athens and disciple of Paul, 2. Saint Denis, the third-century first bishop of Paris and the decapitated, singing martyr of Montmartre, and 3. Pseudo-Dionysius, a sixth-century writer of Greek neoplatonic texts. Hincmar of Rheims, disciple of abbot Hilduin at Saint-Denis, is arguably the first to have expressly and fully related the confusion in the ninth century, at which time this belief was not yet fully fostered. In the twelfth century, however, despite reports of Peter Abelard's critical examination of the truthfulness of Saint Denis' identification with the Areopagite in his *Historia Calamitatum*, Saint-Denis would fully defend and exploit its apostolic status as a royal abbey by attaining to this conflation. For a full exploration of the matter, see Luscombe, "Denis the Pseudo-Areopagite."

[27] On the subject of Dionysius the Areopagite serving as protector and liege lord for the French court, see Gabrielle M. Spiegel, "The Cult of Saint Denis and Capetian Kingship," *Journal of Medieval History* 1, no. 1 (1975): 43–69.

[28] When the count of Vexin at one point retreated to the cloister, French king Philip I (1060-1108), Louis VI's father, took over the count's land which was in fief of Saint-Denis, by which he consequently became first vassal to the abbey. After Louis VI's example as it was set in 1124, it became customary for French kings to take up the banner of Vexin at Saint-Denis before marching to battle. See Barroux, "L'Abbé Suger."

[29] See Laura Hibbard Loomis, "The Oriflamme of France and the War-Cry "Monjoie" in the Twelfth Century," in *Studies in Art and Literature for Belle da Costa Greene*, ed. Dorothy Miner (New Jersey: Princeton University Press, 1954), 67–84; and Spiegel, "The Cult of Saint Denis," 58–9.

> up as he had vowed, as if from his lord, so to speak, he rushed out against the enemy with a small band to protect his person, and sent forth a mighty call for all France to follow him.[30]

Especially through the person of Suger, who was abbot of Saint-Denis from 1122 until 1151, and functioned as close confidant of both kings Louis VI and his son Louis VII (1120–80), Saint-Denis developed strong ties to the Capetian monarchy,[31] a relationship which they emphasized to have originated in the Carolingian period. In reality these ties were a twelfth-century construction driven by a Sandionyisian program, for which Suger functioned as political linchpin. It was Suger who instigated the production of royal historiography in the scriptorium of Saint-Denis, amongst which the Latin text of the *Grandes Chroniques de Saint-Denis* which would be finalized around 1250.[32] It was Suger, as well, who undertook the well-known architectural renovations of the basilica's western façade, famously known to celebrate the historical connection between Saint-Denis and the French monarchy.[33] And, finally, it was under Suger that Saint-Denis grew to a popular destination of pilgrimage.[34] Ultimately these strong affiliations amounted to his acceptance of the regency during the Second Crusade in 1147–9. Suger's future successor, Odo of Deuil, would join Louis VII to battle, and prepare an account of the Crusade.[35]

In this context, where Saint-Denis and the Capetians seem to meticulously and sincerely trace back their ancestry to a common Carolingian origin, we encounter other texts that tap into a similar mythology, yet simultaneously testify of a fictional self-awareness. In these texts, the story of Charlemagne is so playfully reintegrated that it is much harder to believe that the twelfth-century reactualization of this history was taken seriously at all.

One obvious example is the *Pèlerinage de Charlemagne*, an Old French poem that was presumably written by a monk at Saint-Denis (and if not, a writer that was at least intimately familiar to the abbey) supposedly somewhere between 1109-50.[36] The text is a witty parody in which Charlemagne's otherwise heroically depicted journey to the

[30] Original text: "Et quoniam beatum Dionisium specialem patronum et singularem post Deum regni protectorem et multorum relatione et crebro cognoverat experimento, ad eum festinans [...] Rex autem, vexillum ab altari suscipiens, quod de comitatu Vilcassini, quo ad ecclesiam feodatus est, spectat, votive tanquam a domino suo suscipiens, pauca manu contra hostes, ut sibi provideat, evolat, ut eum tota Francia sequatur potenter invitat."Original Latin text (with French translation) taken from Sugerius Sancti Dionysii, *Vie de Louis VI le gros*, § 28, 220; English translation in Sugerius Sancti Dionysii, *The Deeds of Louis the Fat*, trans. Richard Cusimano and John Moorhead (Washington D.C.: Catholic University of America Press, 1992), 24.

[31] Eric Bournazel, "Suger and the Capetians," in *Abbot Suger and Saint-Denis. A Symposium*, ed. Paula Lieber Gerson (New York: The Metropolitan Museum of Art, 1986), 55–72.

[32] Naus, *Constructing Kingship*, 62.

[33] Ibid.

[34] Kathryn Funderburg, "Abbot Suger's St. Denis and the Cult of Relics," *The Expositor: A Journal of Undergraduate Research in the Humanities* 4 (2016): 40–7.

[35] Odo de Deogilo, *De profectione Ludovici VII in Orientem: The Journey of Louis VII to the East*, ed. and trans. Virginia Gingerick Berry, Records of Civilization, Sources and Studies (New York: Columbia University Press, 1948).

[36] Edited and translated in Glyn S. Burgess and Anne Elizabeth Cobby, eds. and trans., *The Pilgrimage of Charlemagne*

East is triggered by a marital trifle. The farce significantly takes its departure in the church of Saint-Denis, where in front of all Frankish nobility Charlemagne puts on his crown and asks his queen if she has ever seen a king more dashing than himself, upon which she naively makes the unfortunate lapsus to answer her astounded husband's question with a positive: Hugo the Strong, emperor of Constantinople, is indeed much better-looking. The elusive poem has been interpreted as meaning to ridicule the Second Crusade's failure (1147–9), in which Charlemagne and his wife should really be taken to represent Louis VII and Eleanor of Aquitaine. Some consider it a 'violent' satire to ridicule the influx of relics in the eleventh and twelfth century, and a joyful comedy to establish the French king's authority to the Eastern emperor.[37] In one and the same breath, the story can be read as undermining Saint-Denis' and the Capetians' attempt to legitimize their undertakings in the Holy Land, especially by aligning these endeavours with those of an egocentric and vain Charlemagne whose sincerity and authenticity is questioned throughout.

If this text was indeed —as has sometimes been presumed— recited on occasion of the aforementioned fair of Lendit (the annual celebration of the relics which Charlemagne had brought from the East to Saint-Denis),[38] this is a telling example of just how much credence a medieval audience gave to Saint-Denis's concoctions. One should remember how Guibert of Nogent, in his *De sanctis et eorum pigneribus*, had already criticized Saint-Denis for resorting to forgery and falsification, when they had allegedly discovered out of nowhere the corpse of the legendary third-century saint Fermin of Amiens.[39] The beauty of the *Pèlerinage de Charlemagne* is that it plays with these rumours, and it apparently does so in a genre that defies clear-cut functions or interpretations. One is never quite sure on deciding which aspects of the poem are subjected to ridicule and which are not.

*(Le Pèlerinage de Charlemagne); Aucassin and Nicolette (Aucassin et Nicolette)*, Garland Library of Medieval Literature, A:47 (New York: Garland Publishing, 1988). For an overview on the connection of the Pèlerinage to Saint-Denis and full references to the poem's intentions and dating, see Alfred Adler, "The *Pelèrinage de Charlemagne* in New Light on Saint-Denis," *Speculum* 22, no. 4 (1947): 550–61.

[37] Walpole, "*Pèlerinage de Charlemagne*," 182–3.

[38] Especially the earliest studies on the *Pèlerinage de Charlemagne* entertained the belief that the poem was connected with the fair of Lendit. Because of its short length, this chanson would have more easily lend itself to recital by a jongleur. This is not a conclusive argument, and there is no strong evidence for it, see Theodor Heinermann, "Zeit und Sinn der Karlsreise," *Zeitschrift für romanische Philologie* 56, no. 4 (1936): 522: "Nun ist es seit G. Paris für die meisten Forscher, vor allem für Bedier, eine unumstößliche Tatsache, eine conditio sine qua non, daß zwischen der Gründung des Lendit 1109 und der Karlsreise die engste Verbindung bestanden habe. Indessen ist in Wirklichkeit aus dem Wortlaut des Gedichtes —ein anderes Kriterium haben wir nicht— absolut kein Zwang zu einer solchen Behauptung zu entnehmen."

[39] "Nec mora in monasterio Sancti Dionisi idem actitatur: parata ab abbate ornatiori capsa dum inde extollitur, dum cum membris caput evolvitur, membranula in martiris naribus reperitur, in qua quod esset Firminus Ambianensis martir expromitur," in Guibertus de Novigento, *De sanctis et eorum pigneribus*, 1, l. 558, 103. Which translates to: "And almost immediately thereafter, the same thing happened at the church of Saint-Denis. There the abbot had a fairly lavish shrine built, and when the martyr was raised up and the head and limbs were uncovered, a strip of parchment was discovered in his nostrils, proclaiming that it was the martyr Fermin of Amiens."Translation taken from Guibertus de Novigento, Monodies *and* On the Relics of Saints. *The Autobiography and a Manifesto of a French Monk from the Time of the Crusades*, trans. Joseph McAlhany and Jay Rubenstein (New York: Penguin Books, 2011), 211.

The most vital document when studying the donation of Charlemagne, and especially when trying to grasp its fictional(ized) nature, is the *Pseudo-Turpin*, which we already came to mention in passing. The Latin chronicle shows a direct, verbal connection to D Kar 286 (see the appendix on p. 366). Unfortunately, instead of offering some grounds upon which D Kar 286 can be dated or contextualized, the *Pseudo-Turpin* has led scholars on a wild goose hunt concerning its authorship, dating, provenance and intentions. The *Pseudo-Turpin* forms part of the larger *Book of Saint James* (*Liber sancti Jacobi*) contained within the *Codex Calixtinus*. Allegedly more than 200 manuscripts of the text exist, and determining which of these is the oldest or most reliable has proven, in retrospect of the various editions with drastically divergent conclusions,[40] a nearly impossible task. The *Jacobus* is indeed one of the best known and most discussed text collections of the twelfth century, and its authorship (was it written by an individual author or by different pens?) is still under much dispute. The direct verbal connection to D Kar 286 has led some scholars to believe that the *Pseudo-Turpin* was composed and/or compiled at Saint-Denis, but there is in fact scant evidence to support this claim.[41] In a drastic but nonetheless important study, Christopher Hohler has argued that the *Pseudo-Turpin*, and especially its passage of the donation of Charlemagne in chapter 22, was not taken seriously for a second. The latter states that the privilege is "absurd nonsense," and that "anyone who imagines that such a document would have enhanced the standing of the abbey, or been taken seriously in any way, is transposing the Paris of John of Salisbury and Adam of St. Victor into the terms of comic opera."[42] According to Hohler, the *Pseudo-Turpin* has all the characteristics of a pedagogical, witty school text, in all likelihood originating from Paris. Its implications for reality are therefore restricted to "some sort of charade," meaning that the donation, which requests funding for the east end of the abbey church, was part of a performance of some sort, "at the end of which the hat would be passed around."[43]

The charter's close connection to the chronicle *Pseudo-Turpin* (the verbal echoes are indicated in detail on p. 366) could nourish the belief that these texts have the same

[40] The difficulty of identifying the text became undeliberately and somewhat embarrassingly obvious by the time of the *Pseudo-Turpin*'s very first conception in print. In the late 1930s, two editions of the *Pseudo-Turpin* were edited and published almost simultaneously by two scholars, Hamilton M. Smyser and Cyril Meredith-Jones (cited in n. 10), who built up divergent stemmas of the text's transmission and reconstructed a different text based on a different archetypal manuscript, see Frederick Maurice Powicke, "Reviewed Work: *The Pseudo-Turpin, edited from Bibliothèque Nationale, Fonds Latin, MS. 17656 by H. M. Smyser*," *Speculum* 13, no. 3 (1938): see 365: "These learned editors differ from each other on almost every important point. Each uses as his basis a text to which the other barely refers [...] Each editor has read widely for the purposes of criticism and annotation, but in directions which are curiously divergent. [...] Neither editor has been able so to widen his horizon as to take full account of a fundamental problem —the relation of *pseudo-Turpin* to the whole historical movement, which had its centre in Saint-Denis, and which had as its object the appropriation of the Carolingian legend."

[41] "Few scholars seriously suggest that the Turpin was composed at Saint-Denis," see Brown, "Saint-Denis and the Turpin Legend," 52.

[42] See Hohler, "A Note on *Jacobus*," 37.

[43] See ibid.

origin: Saint-Denis. This is unclear, and most evidence points to the contrary. Elizabeth Brown has shown that during the twelfth century the history of Saint-Denis and its conflict of interest with the Capetian dynasty is (at least from a material perspective) conspicuously detached from the story of the *Pseudo-Turpin* and its transmission.[44] The texts —D Kar 286 and *Pseudo-Turpin*— are written in a different style and seem to have been conceived in different contexts. This may suggest that the *Pseudo-Turpin* originated from a different nexus of stories, and that only when those stories reached Paris, the story of Charlemagne's donation to Saint-Denis would have been integrated in the cycle. Precisely this matter —which text influenced which text and at what time— is under much dispute, and remains difficult to ascertain. What is nonetheless beyond dispute is that one of these two texts must have branched out from the other.[45] This raises the yet unresolved question whether the writer of the passage in the *Pseudo-Turpin* has adapted the forged diploma of Saint-Denis to conform to a *chanson de geste*'s epic style, or whether D Kar 286 has taken the *Pseudo-Turpin*'s material and embellished it with veracious detail and diplomatic code so that it came to resemble a truthful, legal document.

All of the above is relevant for the present study, as it demonstrates that the donation of Charlemagne lingers in a grey zone that —under the modern lens of a scholar or reader— either clouds our ability to make sense of this text, or else predefines from the very outset the criteria by which it will be judged or interpreted (either as a legal document, or as a fictional text). The fact that the donation avoids a clear-cut categorization of being either a *fictio* or a *lex* through a tension of its form and content, presents its most delusive aspect. Consequently, we are presented with an exemplification of the perfect paradox that a thinking human being's ability to understand a text, i.e. to read it closely and deduce meaning from that text, is also that researcher's main encumbrance in matters of authorship attribution. Hence: our 'understanding' of the text prevents us from understanding it.

## 7.3 The Dating, Authorship, Purpose and Tradition of D Kar 286

### 7.3.1 A Tradition of Authenticity Disputes

The preceding paragraphs sketch the difficult context in which to understand D Kar 286. With a cleverly amalgamated patron, a banner that was (probably inaccurately) proclaimed to be Charlemagne's *Oriflamme*, and an embellished affiliation to the Car-

[44] Brown, "Saint-Denis and the Turpin Legend," 52–5.

[45] A point also made in Groten, "Die Urkunde Karls des Großen," 18–9: "Eine von D 286 unabhangige Entstehung des *Pseudo-Turpin* textes ist schwer vorstellbar, denn woher sollte der Verfasser der Chronik den Gedanken genommen haben, eine Konigskronung dürfe nicht ohne Beratung mit dem Abt von St.-Denis erfolgen?"

olingian dynasty, it becomes undeniably clear that Saint-Denis went through great pains in fabricating its history. In an attempt to unmask the responsible forgers and deconstruct the myths, one can generally discern two models underlying the argumentations of researchers when dating or attributing D Kar 286. Some scholars depart from an assertion that they can somehow discern the charter's historical purpose. They contextualize the text by explaining to the reader its 'true' meaning and impact. Then they conjecture a date. Others do the opposite. They draw *termini post* and *ante quem* in the sand by demonstrating the textual links with other charters or texts whose date of composition are known, and gradually accomplish a stemma. Once the date established, the charter is attributed to whichever likely author alive at that time, and the charter's true historical purpose is revealed through the author. Three possible candidates have already been raised in secondary scholarship: Suger, Odo and Jacques Doublet (who published its transcription in the seventeenth century). One candidate who has been left out of the picture is William of Saint-Denis, Suger's closest secretary. In what follows, we will survey these candidates one by one, and consider some of those arguments that can be raised against and in favour.

### 7.3.2 Suger of Saint-Denis

As hinted at a couple of times by now, one candidate author is abbot Suger of Saint-Denis. In his younger days, Suger provided for himself an intimate familiarity with the archive of Saint-Denis and the muniments kept there. His work and immersion within the *chartrier* of Saint-Denis must have initiated him to the language and formulae of the old Merovingian and Carolingian diplomas. He was ordered by his predecessor, abbot Adam, to order, classify and scrutinize the archive's charters and, indeed, judge them by their value and authenticity. This was, according to Suger himself in the *De administratione*, a pungent concern, since the abbey was plagued by extortions through falsifiers.[46] However, as Jens Clausen has demonstrated, there is ample reason to suspect that Suger was not reluctant to fight fire with fire, and falsified diplomas himself in order to set records straight.[47] In 1129, in the seventh year of his abbacy, he appears to have been responsible for producing a dubious charter, allegedly issued by Louis I, by which he was able to gain possession of the monastery of Argenteuil.[48] This, in combination with allegations of the nuns' scandalous misconduct, gave him the leverage to expel the nuns of Argenteuil —amongst whom Heloise— and acquire

[46] For Adam of Saint-Denis, see Michel Bur, *Suger. Abbé de Saint-Denis, régent de France* (Paris: Librairie Académique Perrin, 1991), 38.

[47] See, for instance, Jens Peter Clausen, "Suger, faussaire de chartes," in *Suger en question. Regards croisés sur Saint-Denis*, ed. Rolf Große, Pariser Historische Studien 68 (München: Oldenbourg, 2004), 109–16; Naus also drew attention to how Suger would cleverly manipulate the timing of important events so as to magnify their impact. Suger appears to have done so in his *Vita*, for instance, where Louis VI's return of the royal regalia in 1120 coincides with the latter's donation of privileges in 1124, granted because of Saint Denis's protection against the German emperor Henry VI. See Naus, *Constructing Kingship*, 65–6.

[48] Waldman, "Abbot Suger."

the dominions for Saint-Denis:

> When in the very impressionable age of my adolescence, I used to leaf through the ancient charters of the abbey's possessions kept in its chests, and since, because of the dishonest deeds of many persons guilty of fraud, I had to consult our collections of immunities, repeatedly there would fall into my hands the foundation charter of the monastery of Argenteuil given by Hermanric and Numma his wife, in which was contained the information that from the days of King Pippin it belonged to the abbey of Saint-Denis. But because of an unfortunate contract, it had been alienated in the time of Charlemagne, his son.[49]

It will be interesting here to ascertain whether or not there are stylistic affinities between the charter for Argenteuil —likely forged by Suger— and the donation of Charlemagne. Aside from suchlike suspicions, it moreover generally fits in well with Suger's programmatic ambitions for Saint-Denis to forge close ties to the Capetians and to trace back the abbey's joint history with that dynasty to the Carolingian line. As already demonstrated higher up through the passage of the Life of Louis the Fat, Suger would recurrently visit themes explored in the donation, such as the French king traditionally paying tribute to Saint-Denis' patron Dionysius, and the taking up of the red banner of Vexin as a sign of the latter's vassality. Suger's initiation to the royal circle of Louis VI and Louis VII has been argued to be so extensive that his Latin style is detectable in the royal charters as soon as his arrival in 1118.[50] In the donation charter, Charlemagne asks of all bystanders to contribute a sum of four bezants to the augmented construction of the church of Dagobert (i.e. Saint-Denis) "up to the cross." This can be linked to Suger's ambitious restoration project of the Saint-Denis church abbey, and the gold-sheathed Great Cross which he placed before the chevet of the church choir.[51]

[49] "Cum etate docibili adolescentiae meae antiquas armarii possessionum revolverem cartas, et immunitatum biblos propter multorum calumniatorum improbitates frequentarem, crebro manibus occurrebat de cenobio Argentoilensi fundationis carta ab Hermenrico et conjuge ejus Numma, in qua continebatur quod a tempore Pipini regis beati Dyonisii abbatia extiterat. Sed quadam occasione contractus incommodi, in tempore Karoli Magni filii ejus, alienata fuerat," see Sugerius Sancti Dionysii, *Gesta Sugerii abbatis (De administratione)*, in *Suger. Oeuvres*, ed. and trans. Françoise Gasparri, vol. 1, Les classiques de l'histoire de France au Moyen Age 37 (Paris: Les Belles Lettres, 1996), §3, 64. Translation taken from Waldman, "Abbot Suger," 242.

[50] Dufour, *Actes de Louis VI*, 59–61.

[51] See the phrase "quatuor singulis annis aureos persoluant nummos pro illius augmento ab edificio Dagoberti regis excellentissimi usque ad crucifixum"in D Kar 286. The last three words, "up to somewhere at the crucifixion,"are somewhat puzzling. They might refer to a cross planted somewhere on the terrain of the abbey church, see Gasparri, "La politique de l'abbé Suger," 241. On the other hand, the "crucifixion" could be taken to refer to the 6 feet-tall gold-sheathed Great Cross which Suger erected in the east end of the abbey church between 1145 and 1147, which came in replacement of a seventh-century cross forged by the confidant of Dagobert, the renowned Saint Eligius (Saint Éloi). I am assuming —although I cannot be certain from his text— that this is the underlying logic behind Hohler's assertion that the charter's "appeal is specifically for the construction of the east end of the church (1140-43 therefore)," see Hohler, "A Note on Jacobus," 37. For Suger's commissioning of the cross, whose splendour is described elaborately in his *De administratione*, §32, see Verdier, "What do we know of the Great Cross of Suger in Saint-Denis," *Gesta* 9, no. 2 (1970):12–5; and Verdier, "La grande croix de l'abbé Suger à Saint-Denis," *Cahiers de civilisation médiévale* 13, no. 49 (1970): 1–31.

### 7.3.3 Odo of Deuil

Odo of Deuil was a young protégé of Suger who acceded as abbot of Saint-Denis after the latter's death in 1151. Little is known of Odo's life, until Suger commissioned him to join king Louis VII's battle in the East as of 1147, with the assignment of recording the king's conquests during the Second Crusade. It is quite possible that at this point Odo's succession had already been decided upon, and that he was meanwhile trained within the monastic administration of Saint-Denis —as Suger would have done under Adam.[52] Odo's account of the Second Crusade, the *De profectione Ludovici VII in Orientem*, is the only work we have transmitted by him.[53] It has been argued that the text was originally meant to have served as exemplar for Suger's fragmentary *Vita* of the king, or was even intended to become an integral part of it.[54] The close relationship between Suger and Odo, the monk's administrative role in the scriptorium of Saint-Denis, and the abbot's sufficient confidence in Odo's capacities to the extent that he integrated the latter's works into his own, raises the question if Odo was indeed not on many occasions Suger's secretary.

There have been scholars, most notably Co Van de Kieft in 1958,[55] who have suspected Odo of Deuil as the better candidate for having forged the donation charter. Van de Kieft's main argument is based on D Kar 286's dating between 1156 and 1165, but his intricate arguments and timelines, deduced from a chain of related charters and texts for which he sets *termini ante* and *post quem*, often appear designed to stitch up an argument that is —as Eric Bournazel did not fail to notice—[56] far more personal. For Van de Kieft, only a morally inferior figure than Suger could have been responsible for such a flagrant forgery. In so doing, Van de Kieft takes the donation of Charlemagne very seriously, and treats it as a fraudulent legislative document instrumental to deceitful and/or criminal intention.

If one thinks back of how Tout had characterized two main "peccadillos" of the Middle Ages, namely murder and lies,[57] Odo of Deuil's reputation was unfortunately blemished twice, since the abbot has been suspected guilty of both. He was charged of

[52] Berry, *De profectione Ludovici VII in Orientem*, xv.

[53] Odo de Deogilo, *De profectione Ludovici VII*.

[54] "The resulting text, then, was intended to form the basis for much of Suger's planned biography of Louis VII," see Naus, *Constructing Kingship*, 79; Suger's *Vita* of Louis VII was left unfinished, although it would later become the starting point of the *Historia gloriosi regis Ludovici*, composed by a monk in Saint-Germain-des-Prés around 1171–3, see Auguste Molinier, "1848. *Historia gloriosi regis Ludovici*," in *Époque féodale, les Capétiens jusqu'en 1180*. Les Sources de l'histoire de France - Des origines aux guerres d'Italie (1494). 2 (Paris: A. Picard et fils, 1902), 183.

[55] Van de Kieft, "Deux diplômes faux."

[56] Bournazel, "Suger and the Capetians," 62: "C. Van de Kieft, on the other hand, places it between 1156 and 1165, during the abbacy of Odo of Deuil. He connects it with the forging, slightly earlier, of another false charter of Charlemagne [...], created to affirm the rights of Saint Denis over certain churches in Berry. Van de Kieft's reasoning is based on internal criticism of the two acts, and also on moral considerations relating to Suger's character. Reminding the reader of Suger's stature as an "homme probe et intègre," he peremptorily concludes that "such a man would never indulge in the forging of false acts." But is this absolutely certain?"

[57] I am referring to Tout's quote given earlier in this thesis on p. 51. For Tout's piece on medieval forgers, see Tout, "Mediaeval Forgers and Forgeries," Tout uses the term 'peccadillo' on 117.

theft and even of murder by some of the Sandionysian monks, and as a consequence had to sedate a true uprising in the first months of his abbacy at Saint-Denis. The situation allegedly escalated to the extent that Odo was forced to call on the authority of Bernard of Clairvaux to defend his innocence.[58] It is especially this type of *ad hominem*-argumentation, which confirms historical rumours and deflects suspicions away from Suger, that Van de Kieft seems to take on. His arguments come with the presupposition that 'forging' in the Middle Ages was always and without question a criminal act that could be associated with the criminal: "It is none the less true that Odo did not equal his illustrious predecessor in probity and moderation; he certainly was not a man to back down from fabricating charters to achieve his goal."[59] Van de Kieft's arguments for making the attribution therefore become questionable.

Nevertheless, it would be unfair to set down Van de Kieft's argumentation as based on personal bias alone. If we disregard the rumours of turmoil under his abbacy, other reasons why Odo makes a suited candidate arise. One of these reasons is that Odo's career had originally started as prior in Chapelle-Aude from 1135 onwards. Chapelle-Aude was located in the diocese of Bourges (in Bas-Berry), and originally founded by Saint-Denis halfway the eleventh century (c. 1059/60).[60] Much like Saint-Denis, the priory had frequently been engaged in fierce arguments over the possession of parish churches in their environment, and there is evidence to demonstrate that in order to lay a claim on these possessions the monks of Chapelle-Aude resorted to forging charters which purported to have been issued by Merovingian or Carolingian authorities.[61] In an intricate study of interrelated charters, Van de Kieft sees returning motives from the forgeries of Chapelle-Aude in later charters forged at Saint-Denis. That D Kar 286 was written during the second half of the twelfth century is further supported by the fact that it might have been a reaction to Frederick Barbarossa's appropriation of the emperor's history resulting in Charlemagne's canonization (Odo died in 1162), or possibly to another false (and insecurely dated) charter, forged by a cleric in Aachen, in which the Aachen church is —much akin to Saint-Denis in the donation— elected "caput omnium civitatum et provinciarum Gallie."[62]

[58] Bernardus Claraevallensis, Epp. 285, 286 and 287, *SBO* 8:200–2.

[59] My translation. "Il n'en est pas moins vrai, qu'Eudes n'égalait pas son illustre prédécesseur en probité et en modération; il n'était certainement pas homme à reculer devant la fabrication de diplômes faux pour atteindre son but." Van de Kieft, "Deux diplômes faux," 421–36.

[60] Émile Chénon, *Histoire et coutumes du prieuré de la Chapelle-Aude* (Paris: Librairie de la société du recueil Sirey, 1915), 15.

[61] Most of them are found published in Martial-Alphonse Chazaud, *Fragments du cartulaire de la Chapelle-Aude* (Moulins: Imprimerie de C. Desroisiers, 1860); a discussion on the forgeries in Chapelle-Aude and a list of suspect charters with page references to Chazaud's edited fragments is given in Co Van de Kieft, *Étude sur le chartrier et la seigneurie du prieuré de la Chapelle-Aude (xi^e-xiii^e siècle)* (Assen: Van Gorcum & Comp. N.V., 1960), 38–41.

[62] Mühlbacher, who posthumously published D Kar 286 for the *MGH*, believed that the diploma was fabricated around

### 7.3.4 William of Saint-Denis

Possibly in the ranks of the faction against Odo of Deuil, although the reasons of why he fell in disgrace with Odo are far from certain, was William of Saint-Denis.[63] William was Suger's biographer, librarian and closest secretary (Suger was encircled by at least five Williams). We know very little of William of Saint-Denis except for his writings, and from these writings so much can be inferred as to conclude that William was well-educated and skilled in Latin, as is to be expected from Suger's right hand. His *Apologetic Dialogue*, dedicated to Odo as a piece soliciting amnesty after the latter had banished him from the abbey to the distant priory of Vaux, has been called a "humanist master piece."[64] In the seventeenth century, Charles d'Auteuil concluded from a comparative analysis between Suger's *De administratione* and his *De rege Ludovico* that William and not Suger was responsible for having written the latter work, a thesis which was —however— soon rejected.[65] Nevertheless, there is validity to the question if a writer of such great skill as William played a part in Suger's oeuvre as it is transmitted today. The secretary never surfaced as candidate forger for the donation to our current interest, which is of course defensible in consideration of the fact that as an authority, William would have had little to gain from forging the donation. His stylistic influence in the charter could nevertheless provide additional evidence for dating the donation, as well as yield a practical new insight into the workings of Suger's chancery.

### 7.3.5 Jacques Doublet

Brown has raised the suggestion that the seventeenth-century Sandionysian monk and editor of the charter, Jacques Doublet, could well have been the actual forger.[66] Unfortunately Doublet wrote his *Histoire de l'abbaye de S. Denys* in French, wherefore I cannot prove nor disprove here Brown's hypothesis by comparing D Kar 286 with samples of Doublet's Latin. Brown's argumentation is primarily based on documentary evidence that the *Pseudo-Turpin* and the chapters it contains on Charlemagne's

December 1165, linking its composition with the canonization of Charlemagne: "Allem Anschein nach kannte der Fälscher die kurz nach 1165 geschriebene legendenhafte *Vita Karoli Magni*, deren Verfasser auch zu St. Denis in näheren Beziehungen stand," see Mühlbacher, *Urkunden*, 428; for the false Aachen charter, see Hugo Loersch, "Das falsche Diplom Karls des Grossen und Friedrichs I. Privileg für Aachen vom 8. Januar 1166," in *Die Legende Karls des Grossen im 11. und 12. Jahrhundert. Mit einem Anhang über Urkunden Karls des Grossen und Friechichs I. für Aachen von Hugo Loersch.* Publikationen der Gesellschaft für Rheinische Geschichtskunde 7 (Leipzig: Verlag von Duncker & Humblot, 1890), 149–215; this dating has been contested by Grauert, "Novitätenschau."

[63] Possibly there existed some envy between Odo and William as two close confidants of Suger. Be that as it may, some event or rivalry between the two had incited Odo of Deuil to force William to stay some time in Saint-Denis-en-Vaux, closeby Châtellerault in Poitou. William stayed in Vaux, as he says himself, "in solitude" and "long-during silence," despite the efforts of monks in Saint-Denis to establish his return. See André Wilmart, "Le dialogue apologétique de moine Guillaume, biographe de Suger (*)," *Revue Mabillon* 32 (1942): 80–118.

[64] Glaser, "Wilhelm von Saint-Denis."

[65] Ibid., 258.

[66] Brown, "Saint-Denis and the Turpin Legend."

donation can be brought in connection to the abbey only from the thirteenth, fourteenth and fifteenth century onward, which is when the chronicle gained popularity through its many vernacular translations and was gradually accepted in Saint-Denis despite skepticism toward its preposterous claims. Brown's hypothesis largely rests on the verbal overlap and dependencies between the charter and that of post-twelfth-century translations (which she argues to be more striking than those in the 'original' Latin *Pseudo-Turpin*). She also argues that only after the twelfth century there is evidence that the custom of donating the four bezants had become installed. Only throughout later centuries the abbey increasingly grew fond of the myth, and celebrated the Turpin's contents until they reached a point of no return. Only later in the late fifteenth century, when skepticism was raised against the legitimacy of the claims, the abbey was forced to defend itself. In order to uphold the reputation of Saint-Denis, the monk Jacques Doublet forged the donation charter, which later became a red herring for scholars today.

All in all I cannot but conclude that if one wants to agree to Brown's arguments one needs to accept a considerable number of premises, not all of which equally convincing. Brown attaches much weight to material transmission, the presence of verbal reminiscences in contemporary and later adaptations / translations, and eye witness accounts of originals now lost. As she herself admits, there is no evidence to contradict the strong indications that François de Belleforest (1530–83), himself skeptical of Carolingian myths, laid eyes on the donation of Charlemagne in the late sixteenth century —i.e. long before Doublet's *Histoire* was published. Brown consequently hypothesizes that de Belleforest had been deceived by another forgery or had only been indirectly informed on its contents. Although Brown's research is thorough and well-documented, it results in a too speculative story with too many machinations and shaky arguments to be truly convincing. The straightforward and least far-fetched explanation still seems most credible to me: here is a medieval author at work who knew the archives of Saint-Denis well, operating at a time when forgery was at full sway, and at a time when Saint-Denis's ambition was to forge ties to the Carolingian past.

### 7.3.6 The Corpus —and Some Caveats

The training corpus representing each of our three candidates is assembled in table 7.1. Some caveats should be formulated. The first of these appertains to the genre of the charter. Charters have a specific and formulaic style, one by which many authors could be trained. Therefore they are both the result of and basis for collaborative forms of authorship. Aspects that typify charters are seldom linked to individual stylistic traits. More often one finds them categorized according to transcendent 'stylistic' typicalities proper to the period, region and chancery (i.e. an administrative *group* of writers, schooled in a particular manner). In this light it can often be observed that dating charters seems more pertinent to medievalists than attributing them. Also computational

research into charters has often restricted its focus to the evolution of palaeographic principles or the recognition of handwritten aspects of style rather than picking up on the lexical or syntactic elements that appertain to the style of the scribe behind the diploma.[67] This experiment, in other words, therefore also sidetracks into an additional subquestion: can charters be subject to individual contributions on a stylistic level? To a large extent this harks back to the discussions within the previous chapters, where genre has been shown to occasionally subdue the voice of the individual. Suger, Odo, and William, have been observed to have written much like each other, which can be explained by their having been educated in comparable curricula. Are these writers distinguishable, if their literary upbringing had been this resemblant, especially in the genre of charter? This is an interesting question, as scholars such as Françoise Gasparri[68] and Jean Dufour[69] seem to have been able to spot recognisable features within Suger's style and link them to his person nevertheless.

The second caveat appertains to a discrepancy between the training and test data, and the class imbalances within the corpus as a whole. In the successive paragraphs, our three candidates' stylistic profiles will be set up by means of texts that are wholly different from charters. Suger's letters are unlike William's apologetic dialogue, and Odo's history of the crusade —although some argue its style is much better appreciated when considering it a letter—[70] will have a very different tone than Suger's *De administratione*. Another practical problem within the corpus is the length of the texts, and the 'scarcity' by which some authors are represented. For William of Saint-Denis we lack, for instance, historiographical material. One can also look at the word count column and deduce that Suger is far better represented in volume than the two other candidates (his corpus is more than twice the size of the other two candidates). In addition to this imbalance, this diverse training data containing narrative material is tested against very short charters. The donation of Charlemagne is only 710 words long, and this includes the appended location, date and long list of signed witnesses at the end of the charter (112 words), which cannot possibly contain stylistic information. Therefore, in the next few experiments, we will resort to a very short sample length of only some 500 words. In light of all of these problems, it becomes all the more urgent to be well equipped statistically, and to place the necessary question marks next to any attribution that may follow.

[67] See, for instance, the computer-based dating of medieval charters (1300–1550) from the Medieval Dutch language area, in Sheng He et al., "Towards Style-Based Dating of Historical Documents," in *Proceedings of 14th International Conference on Frontiers in Handwriting Recognition* (2014).

[68] Françoise Gasparri, "Le latin de Suger, abbé de Saint-Denis (1081–1151)," in *Les historiens et le latin médiéval*, ed. Monique Goullet and Michel Parisse, Histoire ancienne et médiévale 63 (Paris: Publications de la Sorbonne, 2001), 177–193.

[69] Dufour, *Actes de Louis VI*, 59–61.

[70] Caron Ann Cioffi, "The Epistolary Style of Odo of Deuil in his 'De Profectione Ludovici VII in Orientem'," *Mittellateinisches Jahrbuch* 23 (1988): 81.

| Author | Title (or *incipit*) | Ed. | Length |
|---|---|---|---|
| Suger of Saint-Denis | *De consecratione* | F. Gasparri | 4,873 *w* |
| | *De rege Ludovico* | Ibid. | 1,929 *w* |
| | *Epistolae* | Ibid. | 6,615 *w* |
| | *Fragmentum vitae Ludovici Iunioris* | Ibid. | 1,405 *w* |
| | *Gesta Sugerii* (*De admin.*) | Ibid. | 9,998 *w* |
| | *Vita Ludovici grossi* | H. Waquet | 23,539 *w* |
| Odo of Deuil | *De profectione Ludovici VII in Orientem* | V. Berry | 15,167 *w* |
| William of Saint-Denis | *Ad quosdam ex suis comonachis* | E.-R. Labande | 884 *w* |
| | *De morte Sugerii* | A. Lecoy de La Marche | 1,429 *w* |
| | *Dyalogum apologiticum* | A. Wilmart | 12,435 *w* |
| | *Vita Sugerii* | F. Gasparri | 5,399 *w* |

Table 7.1: Corpus for candidate authors of D Kar 286: Suger of Saint-Denis, Odo of Deuil and William of Saint-Denis. Full bibliographical references are to be found in the primary sources given on pp. 377ff.

## 7.4 Experimental Set-Up and Preparations

### 7.4.1 Developing a Model: Suger's and Odo's Charters

This section first seeks to ascertain whether or not Suger's and Odo's own charters can be successfully affiliated to their own stylistic profiles by training a classifier on the corpus described above (unfortunately, there are no charters which can be securely attributed to secretary William's hand). Until now, these charters had been left out of the training data set (see table 7.1), as their inclusion would entail bias in favour of authors who have charters included in their training material, which is mostly true for Suger (of whom we have plenty), and least true for William (of whom we have none). The charters, then, can function as our 'development data.'[71] We can test whether or not the training material —of such a different genre than charters— is indeed sufficient enough before proceeding to the attribution of the anonymous charter D Kar 286. This preliminary step likewise tests the assumption whether or not charters are susceptible to the stylistic influence of the 'authority' that issued them. Suger has left us a corpus of sixteen charters, nowadays published in Gasparri's edition of Suger's complete oeuvre.[72] Aside from these 16, 8 of Louis VI's royal charters are generally believed to have been written by Suger under his administration, more specifically those charters containing privileges granted to Saint-Denis.[73] Some 3 charters presumably transmitted under Odo's administration are edited from the *Cartulaire blanc* —only

[71] The ideas of training, test and development data are more fully expounded upon in the appendix on pp. 337ff.

[72] Sugerius Sancti Dionysii, *Chartes*, in *Suger. Oeuvres*, ed. and trans. Françoise Gasparri, vol. 2, Les classiques de l'histoire de France au Moyen Age 41 (Paris: Les Belles Lettres, 2001), 156–267.

[73] Dufour, *Actes de Louis VI*, 59–61.

two of which can be securely dated—, and are available in the cartulary's digital edition online.[74]

### 7.4.2 Training SVM Models

The attribution of the charters was first tested in a 'closed' setting. This entailed that a series of SVM classifiers (support vector machine) with different settings were optimized to segregate the respective authors' documents in the training corpus (table 7.1).[75] For each of the three 'favourite candidates' the algorithm draws up stylistic profiles, after which the test data —i.e. the charters of Suger and Odo— were attributed. Note that a 'closed game' entails that the algorithm *always* makes an attribution, which makes it different from the *impostors* method, where a confidence threshold needs to be attained in order for the model to even make an attribution at all. In total the training process yielded some 288 SVM models. 72 of these were retained, since the optimal *c* parameter proved to be *c*=1 for all models. With a wide diversity in average performance, ranging from an average of 49.94% to 96.35%.[76] The settings searched were identical to those in chapter 4 (p. 141), with exception of one additional setting: random sampling was introduced against normal sampling. This means that both normal 'segmented' training samples were trained vs. randomly extracted samples (up to 140 per author, all samples were 500 words long).

Random sampling entails that sentences of the respective candidates were extracted by random selection from their different works, and consequently reintegrated within a 500-word sample.[77] In a recent article, David Hoover has argued that random sampling can be beneficial in cases where the training data is scarce, or where 'intra-author style variations' in the existent data —such as genre and register— could possibly interfere with the dominant authorial signal.[78] Scrambling sentences, then, could be a means of smoothening out whatever minor interferences are brought about by the author's deliberate shifts in tone. In our current scenario, where the diversity in genres is quite high, and where a strong discrepancy between the available training data and test data exists, it seemed worthwhile assessing the performance of this technique.

Table 7.2 yields a ranking of the average performances for the top 5 and lower 5

[74] The *Édition électronique des chartes de l'abbaye de Saint-Denis*. The *Cartulaire blanc* is a medieval compilation containing 2,600 copies of charters from Saint-Denis from the ninth to the thirteenth century. The charters here referred to are numbered as Saint-Martin 3 (1151–8) and 4 (1151–9) and Tremblay 9 (1151-69). Each of these is written in person of Odo. Tremblay 9, however, is dubious, since its contents allow its composition to have taken place after 1162 by Odo's successor, namely Odo III. For more about the online edition's constitution, see Clavaud, "The Digital Edition," the digital edition can be consulted here: `http://saint-denis.enc.sorbonne.fr/` .

[75] For a more detailed explanation of support vector machines (SVM), see chap. 3 on pp. 94ff., and this thesis's appendix on pp. 336ff. For its application to computational stylistics specifically, see Diederich et al., "Authorship Attribution with Support Vector Machines."

[76] These percentages were based on a mean over accuracy, precision and recall, the concepts of which are explained on p. 342.

[77] For more on randomization or random sampling, see pp. 319ff.

[78] David L. Hoover, "The Microanalysis of Style Variation," *Digital Scholarship in the Humanities* 32 (2017): ii17–ii30.

| Rank | Weighting | Feature type | Number | Sampling | CV-method | Score |
|---|---|---|---|---|---|---|
| 1 | normal | function words | *n*=250 | normal | stratified | 0.963498 |
| 2 | tfidf | words | *n*=250 | normal | stratified | 0.963498 |
| 3 | normal | words | *n*=250 | normal | stratified | 0.963498 |
| 4 | tfidf | function words | *n*=250 | normal | stratified | 0.963498 |
| 5 | normal | function words | *n*=175 | normal | stratified | 0.957230 |
| ... | | | | | | |
| 68 | tfidf | function words | *n*=50 | random | KFold | 0.503854 |
| 69 | normal | function words | *n*=100 | random | KFold | 0.503854 |
| 70 | tfidf | 4-grams | *n*=50 | random | KFold | 0.502956 |
| 71 | normal | words | *n*=50 | random | KFold | 0.500263 |
| 72 | normal | 4-grams | *n*=50 | random | KFold | 0.499365 |

Table 7.2: Ranked results of SVM models' performance in distinguishing between the works of Suger of Saint-Denis, Odo of Deuil and William of Saint-Denis. There were 288 models in total, but all models consistently had *c*=1 as optimal parameter, wherefore only 72 are shown. The larger mid-section is not shown, which allows to compare the parameters distinguishing high and low performance.

SVM models in distinguishing between Suger's, Odo's and William's different writing styles. It becomes clear that stratifying the training data is beneficial (as opposed to an equal folding technique), and that more features appears to work favourably rather than fewer, despite the text samples' short length. Character 4-grams are also less likely to yield good distinctions as opposed to lexical items. Perhaps surprisingly, the models in which samples had been randomly assembled did not attain high scores in this scenario. Randomization seemed to confuse the algorithm rather than increase its performance. The PCA plot (fig. 7.1) gives a 2D intuition of how the highest ranking classifier (96.35%) succeeded at setting apart the three authors' works. Note that the actual classifier's decision boundary —which cannot be visualized here— takes into account an additional 248 (largely invisible) dimensions of information in its decision function.

### 7.4.3 First SVM Set-Up

Table 7.3 gives the first results of testing the attribution of Suger's and Odo's charters in a closed setting. Also the attribution results for the royal charters, the charter of Argenteuil[79] and the donation charter of Charlemagne (D Kar 286) are listed. The middle column indicates which author was predicted of having written the charter by the majority of 72 SVM models. The percentage in the right column expresses how large this majority is, and grants an additional figure by which to estimate the general confidence and validity of the prediction. An approximation of how to visualize these results is given in a PCA plot (fig. 7.2), where the test texts were plotted within the 'zones' to which the algorithm assigns them.

[79] This is the aforementioned forged charter of 1129, by which Suger allegedly gained possession of the monastery of Argenteuil, see Waldman, "Abbot Suger."

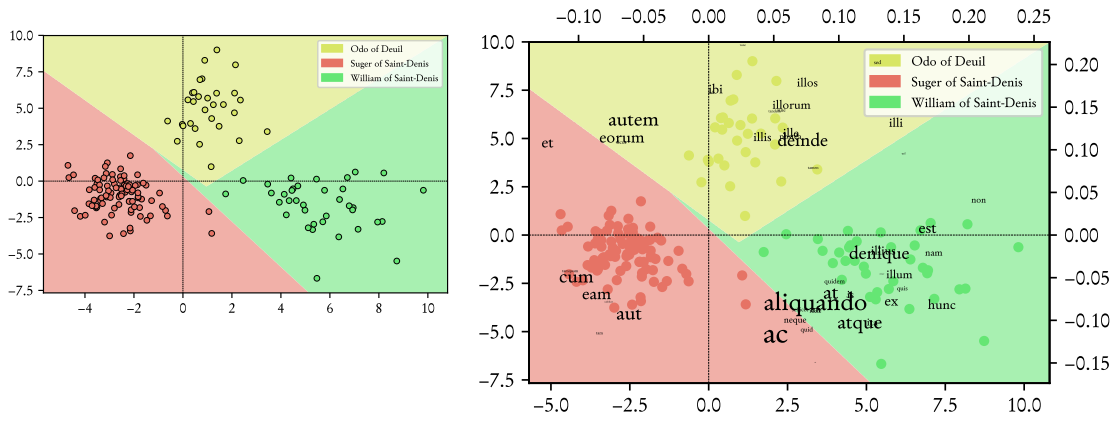



Figure 7.1: PCA plot giving a 2D intuition of the decision boundaries drawn by the best-scoring SVM classifier (whose settings can be consulted in table 7.2). The background colours indicate the respective 'zones' of each of the candidate authors Odo (yellow), Suger (red) and William (green). The right-hand figure shows the 50 (out of 250) loadings. Only the highest ranking loadings on both PC's have been visualized (the font size expresses the weight). Settings: *s–l* = 500 | *type* = most-frequent function words | *n* = 250 (50 most important are visualized) | *vect.* = standard-scaled tfidf-weighted frequencies | *expl. var.* = 8.26%.

If one was hoping to find that charters betray artefacts of personal style, (s)he might foster some optimism in observing that for Suger's charters the model behaves as one would expect. The charters written under Suger's abbacy —as also those he would allegedly have written under king Louis VI's administration— are attributed to him with a high agreement amongst all classifiers. Also the forged charter of Argenteuil clusters on Suger's side. However, the same does not happen for Odo's charters. They, too, are attributed to Suger instead, although the algorithm's insecurity concerning this prediction is expressed by the rather poor confidence scores in the right-hand column (0.69 and 0.53). On closer inspection the results are even further problematized, as the remaining percentage of classifiers judges that Odo's charters were written by secretary William (on average 36.92%), and not in favour of Odo at all (on average 3.47%). That the charters issued under Odo's abbacy (and signed under his name) were written by Suger is obviously impossible; that the algorithm grants second place to William, on the other hand, not necessarily. However, it requires that we assume William was present in Saint-Denis under Odo's abbacy, and active in the scriptorium under Odo's administration, of which we must remain skeptical.[80]

Instead of assuming a direct influence by Suger or William on Odo's charters, a more plausible explanation presents itself in assuming that Odo's charters seek out the company of Suger's documents because the latter's style is completely imbued by diplomatic discourse. That his expertise in historical muniments is reflected in his

[80] Both of these assumptions are problematic considering William's difficult relationship with Odo, and the lack of ev-

| Sample | Prediction | Confidence |
|---|---|---|
| **Suger's charters** | | |
| diploma-1 | Suger of Saint-Denis | 1.00 |
| diploma-2 | Suger of Saint-Denis | 1.00 |
| diploma-6 | Suger of Saint-Denis | 0.89 |
| diploma-8 | Suger of Saint-Denis | 1.00 |
| | Suger of Saint-Denis | 1.00 |
| | Suger of Saint-Denis | 1.00 |
| diploma-11 | Suger of Saint-Denis | 0.99 |
| diploma-12 | Suger of Saint-Denis | 1.00 |
| | Suger of Saint-Denis | 0.86 |
| | Suger of Saint-Denis | 1.00 |
| | Suger of Saint-Denis | 1.00 |
| **Royal charters** | | |
| carta-189 | Suger of Saint-Denis | 1.00 |
| carta-220 | Suger of Saint-Denis | 1.00 |
| carta-281 | Suger of Saint-Denis | 1.00 |
| carta-409 | Suger of Saint-Denis | 0.61 |
| **Odo's charters** | | |
| Odo's charters (1) | Suger of Saint-Denis | 0.69 |
| Odo's charters (2) | Suger of Saint-Denis | 0.53 |
| **Suspect charters** | | |
| Charter of Argenteuil | Suger of Saint-Denis | 1.00 |
| D Kar 286 | Suger of Saint-Denis | 1.00 |

Table 7.3: Attribution results of Suger's charters (for Saint-Denis and under name of Louis VI), Odo's charters, and the suspect charters of Argenteuil and the donation of Charlemagne. 288 SVM models were trained with different settings to distinguish between Suger's, Odo's and William's respective writing styles (p. 141), but all models consistently had $c$=1 as optimal parameter, wherefore only 72 are shown. The prediction made by the majority of classifiers is indicated in the middle column. The right-hand column indicates how large this majority is, thereby yielding a naive kind of confidence score.

writing style has been subject of a special study by Gasparri.[81] It does, indeed, make one wonder to what extent Suger's style has not become an amalgamation of phrases and formulas he allegedly read so frequently in the Saint-Denis archives, a style to which all other charter-like documents are drawn, especially those pertaining to Saint-Denis. But this, at the same time, shows the problematic nature of the methodology,

idence that William ever made it back to Saint-Denis after his expulsion. There is, however, a Latin translation of the panegyric of Saint Denis (originally in Greek), with a preface letter in which a monk named William dedicates the work to abbot Ivo II of Saint-Denis. The preface letter allows to date the translation (which was never edited) to some time around 1162–72. Both Loenertz as Mayr-Harting do not question this attribution and appear confident that the translator is indeed William, former secretary of Suger. The implication is that William was active in Saint-Denis for a considerably longer time after Suger's death, and would have survived Odo in life years. See Raymond Loenertz, "Le panégyrique de S. Denys l'Aréopagite par S. Michel le Syncelle," *Analecta Bollandiana* 68 (1950): 97; and Henry Mayr-Harting, "Odo of Deuil, the Second Crusade and the Monastery of Saint-Denis," in *The Culture of Christendom. Essays in Medieval History in Commemoration of Denis L. T. Bethell*, ed. Marc Anthony Meyer (London: Bloomsbury Publishing, 1993), 239.

[81] "Cette parfaite connaissance du discours diplomatique se rencontre dans tous ses écrits, même au travers de tournures qui lui sont tout à fait personnelles, car son latin est particulier," see Gasparri, "La politique de l'abbé Suger," 236.

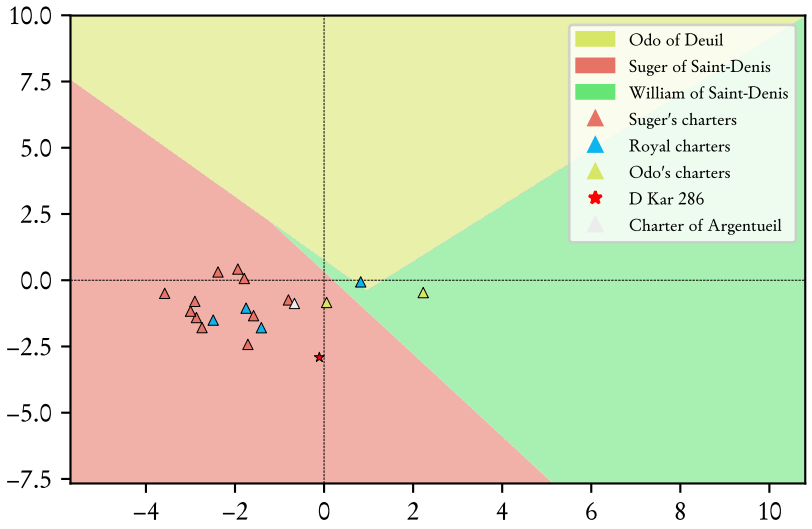



Figure 7.2: Attribution results of table 7.3 visualized in PCA plot. The coloured backgrounds, separated by decision boundaries, indicate the 'zones' as they were demarcated during training (also visible in fig. 7.1).

when genres take the lead in sending out the predominant stylistic signal.

In order to further test the aforementioned assumption —that Suger writes 'in a diplomatic style' clouding our algorithm's performance—, a new experiment was carried out, where all Latin charters within the Saint-Denis *Cartulaire blanc* (± 51,605 *w*)[82] were tested on this same series of 72 SVM models. The results further confirmed the hypothesis above, and raise the warning that the current experiment forms no sufficient evidence to state that D Kar 286 is a charter written by Suger. Nearly all 103 samples taken from this pool of charters was —with an average confidence of 85.47%— found attributed to Suger. In itself, this provides a fascinating observation, as it certainly adds texture to the stylistic profile of Suger, and puts a number to the extent by which the abbot was indebted to his study of charters in the Saint-Denis archives when he was a young pupil. On the other hand, this affiliation is of course problematic on a methodological plain. The diplomatic peculiarities proper to Suger's writing style severely invalidate the reliability of the attribution results. It likewise reinforces the need to depart from a closed setting into a more 'open' setting if one hopes to find a more reliable basis on which to attribute D Kar 286.

### 7.4.4 Second SVM Set-Up

A second set-up seemed advantageous in order to corroborate the reported results. The idea was to lure away the elements that make Suger's style 'diplomatic' to another class, consisting solely of charters. Instead of using the *Cartulaire blanc* as test material

[82] On the *Cartulaire blanc*, and the data's availability online in a digital edition, see n. 74. For more on the edition, see Clavaud, "The Digital Edition."

| Sample | Prediction | Confidence |
|---|---|---|
| **Suger's charters** | | |
| diploma-1 | Suger of Saint-Denis | 1.00 |
| diploma-2 | Suger of Saint-Denis | 0.57 |
| diploma-6 | Suger of Saint-Denis | 0.67 |
| diploma-8 | Suger of Saint-Denis | 0.99 |
| | Suger of Saint-Denis | 1.00 |
| | Suger of Saint-Denis | 1.00 |
| diploma-11 | Suger of Saint-Denis | 0.89 |
| diploma-12 | Suger of Saint-Denis | 1.00 |
| | Suger of Saint-Denis | 0.82 |
| | Suger of Saint-Denis | 1.00 |
| | Suger of Saint-Denis | 1.00 |
| **Royal charters** | | |
| carta-189 | Suger of Saint-Denis | 0.97 |
| carta-220 | Suger of Saint-Denis | 0.97 |
| carta-281 | Suger of Saint-Denis | 0.64 |
| carta-409 | *Cartulaire blanc* | 0.78 |
| **Odo's charters** | | |
| Odo's charters (1) | *Cartulaire blanc* | 0.89 |
| Odo's charters (2) | *Cartulaire blanc* | 0.56 |
| **Suspect charters** | | |
| Charter of Argenteuil | Suger of Saint-Denis | 0.97 |
| D Kar 286 | Suger of Saint-Denis | 0.92 |

Table 7.4: Attribution results of Suger's charters (for Saint-Denis and under name of Louis VI), Odo's charters, and the suspect charters of Argenteuil and the donation of Charlemagne. 288 SVM models were trained with different settings to distinguish between Suger's, Odo's and William's respective writing styles (p. 141), but all models consistently had $c$=1 as optimal parameter, wherefore only 72 are shown. The prediction made by the majority of classifiers is indicated in the middle column. The right-hand column indicates how large this majority is, thereby yielding a naive kind of confidence score.

—as was the case in the previous paragraph—, we involved the charters within the collection as additional training material. This entailed training a second series of SVM models to learn the distinction between Suger and the *Cartulaire blanc* charters (which are of various authorship). Afterwards, the SVM models passed their judgement on the same set of test charters as before, including D Kar 286. The idea was to place the donation charter before a dilemma by forcing it to divide its two 'loyalties.' Either D Kar 286's style affiliates best with a class containing Suger's writings, or else with a class of loosely connected twelfth- and thirteenth-century charters composed in Latin at Saint-Denis. Basically this entails 'trapping' D Kar 286 into revealing where its true loyalties lie: with the author, or with the genre in which Suger so happened to be well-trained. This new set-up yielded a novel batch of results for the same set of charters as in table 7.3.

Having introduced the *Cartulaire blanc* as an additional training class proves fruit-

ful. It manages to place into context the previous hesitant attribution of Odo's charters to Suger. These charters can hardly be brought in relation to both abbots Suger and Odo, and side with the *Cartulaire blanc* class as soon as this option presents itself. This might perhaps suggest that Odo delegated his administration more often to secretaries than Suger, or was far less personally involved in their composition. Suger had increasingly been forced to delegate more toward the end of his life, as Saint-Denis flourished.[83] One of Suger's royal charters (charter 409) also seeks its way out of having to decide between either of the three candidates. Apparently, the stylistic material contained in this royal charter is not convincing enough to be brought in relation to Suger. Most to our interest here is that D Kar 286 was barely tempted to affiliate with the other charters, and confidently remained on Suger's side of the decision boundary. Occasionally, the stylistic affiliation which D Kar 286 appears to have with Suger's writing style was even stronger than that of the charters traditionally attributed to the abbot (especially diplomas 2 and 6 tended to succumb quite a bit to the draw of the *Cartulaire blanc* class). Wen forced to choose between the author and the genre, the attraction towards Suger's style is so powerful that D Kar 286 barely budges. Slowly but surely, Suger's influence on the donation of Charlemagne becomes likelier.

### 7.4.5 *Impostors* Method

Here we perform a final test. In essence, the involvement of an additional class in the foregoing SVM experiment which serves as a kind of 'decoy' to test the attribution's stability is already a first step toward an 'open' game, or the verification problem.[84] The more candidates we introduce, and the less those these new candidates can impress D Kar 286 to change loyalties away from Suger, the likelier the odds become that the attribution to Suger is no coincidence, but results from a deep and consistent stylistic similarity. The most extreme and most developed method which tests this assumption is the *impostors* method, which has by now been explained a number of times.[85] Just as in all the previous *impostors* experiments, a threshold $\sigma^*$ first had to be devised (visualized in fig. 7.3). This threshold will indicate how many times out of $k$=100 times a same and single 500-word text sample of an author is recognized to have indeed been written by that author (<*same-author*> pair), by matching it against the full impostors pool. In the meantime, a feature subset half the size of the total vocabulary is randomly composed $k$ times from the original corpus.

The background corpus consisted of twenty-nine impostors including Suger and the *Cartulaire blanc* class. In the preliminary training rounds, Suger's, Odo's and

[83] Giles Constable, "Suger's Monastic Administration," in *Abbot Suger and Saint-Denis. A Symposium*, ed. Paula Lieber Gerson (New York: The Metropolitan Museum of Art, 1986), 17–32.

[84] I will refrain here from explaining authorship verification in full, as the concept has been made clear in the thesis throughout other case studies, and is more elaborately explained in chapter 3 on pp. 114ff.

[85] The *impostors* method was introduced in the previous case studies, but an insightful survey of its workings can be found at the end of chapter 3 on pp. 117ff. The method was first proposed in Koppel and Winter, "Determining."

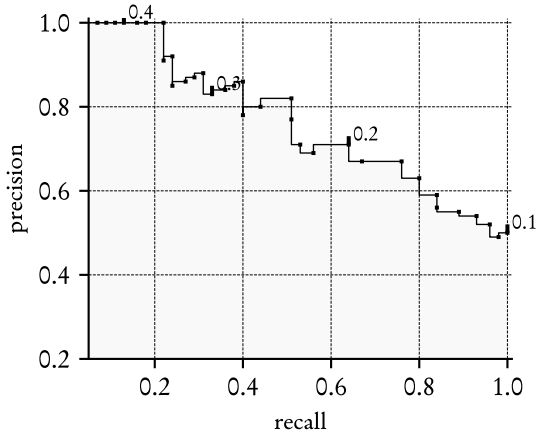



Figure 7.3: PR-curve for the *impostors* method with different $\sigma^*$ thresholds in a preliminary training round for 500-word samples of Suger of Saint-Denis, Odo of Deuil and William of Saint-Denis (tfidf-weighted vectors with standard-scaling). An ideal trade-off between recall and precision is found at a $\sigma^*$ value of 0.23 (see table 7.5 for evaluation metrics in detail).

| | **dev set** | **test set** |
|---|---|---|
| **accuracy** | 0.70 | 0.73 |
| **precision** | 0.82 | 0.73 |
| **recall** | 0.51 | 0.73 |
| **f1** | 0.63 | 0.73 |

Table 7.5: Evaluation metrics for training the *impostors* method on Suger of Saint-Denis, Odo of Deuil and William of Saint-Denis.

William's own text samples were reliably attributed from a $\sigma^*$ threshold of 0.23 onward. This threshold yielded a favourable trade-off at the model's maximal accuracy of 70% (on the development set) and 73% (on the test set). When running this same model on charter D Kar 286, Suger outrivalled all other candidates, with his name surfacing as top candidate 62 times out of 100 (62%). Most of the other attributions were lost on the randomly sampled charters from the *Cartulaire blanc* class, with a score of 19%. William and Odo both stranded at 0%.

It should be noted here again that the *impostors* method as a verification method is far more severe than the normal 'closed' attribution problems (such as those with SVM above), and penalizes poorer stylistic affinities by refraining from classification. Interestingly, both Suger's and Odo's charters succumb quite a bit to the draw of the impostor documents, not surprisingly by their generic affiliation to the charters in the *Cartulaire blanc*. It is all the more striking, then, that this does not happen for the donation charter, which even affiliates better to Suger than some of the charters which are currently ascribed to him. The suspect charter of Charlemagne scores 39% over the original $\sigma^*$ threshold, and does not budge under the stylistic affinity to other Saint-Denis charters.

| Sample | Prediction | Confidence |
|---|---|---|
| **Suger's charters** | | |
| diploma-1 | — | — |
| diploma-2 | — | — |
| diploma-6 | — | — |
| diploma-8 | — | — |
| | Suger of Saint-Denis | 0.57 |
| | — | — |
| diploma-11 | — | — |
| diploma-12 | Suger of Saint-Denis | 0.93 |
| | — | — |
| | — | — |
| | — | — |
| **Royal charters** | | |
| carta-189 | Suger of Saint-Denis | 0.56 |
| carta-220 | — | — |
| carta-281 | Suger of Saint-Denis | 0.64 |
| carta-409 | Suger of Saint-Denis | 0.56 |
| **Odo's charters** | | |
| Odo's charters (1) | — | — |
| Odo's charters (2) | — | — |
| **Suspect charters** | | |
| Charter of Argenteuil | — | — |
| D Kar 286 | Suger of Saint-Denis | 0.62 |

Table 7.6: Attribution results for the *impostors* method applied to Suger's charters (for Saint-Denis and under name of Louis VI), Odo's charters, and the suspect charters of Argenteuil and the donation of Charlemagne. Prediction labels transgressing the $\sigma^*$ threshold of 0.23 are given in the middle column. If verification was impossible (and stranded under the threshold) the column was left blank, indicated as '—.' The right-hand column indicates the confidence score.

## 7.5 Results and Conclusions

Having now applied the entire battery of statistical techniques to D Kar 286 (summarized in table 7.7), all the evidence strongly suggests that Suger —not Odo nor any secretary— was responsible for the authorship of D Kar 286. This conclusion enables us to revisit the introductory remarks to this chapter, which predominantly tied in with pertinent questions around the tension between 'fact' and 'fiction' in twelfth-century literature. What role can computational authorship attribution have in this debate?

The bad news is that getting a full grasp over the donation of Charlemagne might well be a lost cause. That medieval writers interweave the actual with the fictional is a well-known hermeneutic problem, and becomes even more acute when stumbling upon an anonymous document such as D Kar 286, which is essentially left unexplained and has been devoid of context. Especially in these scenarios it becomes clear just how difficult it is for modern-day readers to get a grip on medieval text. The confusion

| Experiment | Confidence | Attribution |
|---|---|---|
| SVM 1 | 1.0 | Suger |
| SVM 2 | .92 | Suger |
| *Impostors* method | 0.62 (+39%) | Suger |

Table 7.7: Final results for the attribution of D Kar 286. The percentage +39% indicates the degree by which the confidence of the attribution transgressed $\sigma^*$.

perhaps originates from the —not coincidental— convergence of two significant developments: the rise of forgery (as well as its counter-reactions), and the rise of literary fiction. One might find it unfortunate that the respective understanding of both of these important 'movements' is compromised by the existence of the other. Their co-existence, —to the modern-day reader / scholar, at least— blurs the distinction between texts celebrating the fictional versus texts composed with upright deception as aim. Naturally, there is evidence that both types of text existed in the Middle Ages, as that of many hybrid forms filling out the spectrum in between. The risk is that the interpretation of D Kar 286 is likely to become a matter of perspective and background. In other words, the subjectivity of the researcher will organically take the upper hand.

But there is good news too. Although it cannot answer to all of these historical-literary questions, statistical attribution triggers reflection on the broader implications of this subjectivity and ideologically charged motives. It demonstrates that qualitative attributions are liable to change from one perspective on authorship to another, and can be susceptible to intuitive responses. Van de Kieft's attribution to Odo of Deuil, for instance, came with a firm belief that the fabrication of documents was always a criminal act, hence the charter's attribution to a 'criminal' character.[86] The evidence above provokes some thought on the extent by which ideology and theoretical mindsets affect scholars' ideas and decisions. Elizabeth Brown, both in the study of the donation charter as elsewhere, has expressed dismay over that 'fictionality' is a card that can be played to exculpate the forgers and the institution they represent. She refuses to believe that forgery was experienced differently at all in the Middle Ages, stating that there is no formal or documentary evidence supporting the claim that medieval writers "denied the reprehensibility of forgery or were convinced by persuasive justifications that unimpeachable motivations excused the act of forgery."[87] Firmly stating that "forgery is now and was in the Middle Ages a crime,"[88] she criticizes amongst others Constable's relativist interpretations of text theft and text fabrication in the Middle Ages.[89] Between the lines, Brown seems to suggest that this lack of

[86] Van de Kieft, "Deux diplômes faux."

[87] Brown, "*Falsitas pia sive reprehensibilis*," 106. There is much to be said for this argument, not in the least when thinking of Guibert of Nogent's *De sanctis et eorum pigneribus*.

[88] Ibid., 101.

[89] "Following the logic of his argument concerning forgery, Constable holds that the copying of sources which took place

criticism simply provides a convenient exoneration of the Church institution.[90] She appears to have seen a partisanship of the like for D Kar 286 when suggesting —be it carefully— that the seventeenth-century Dionysian monk Jacques Doublet could well be the actual forger of the charter.[91] Due to lack of samples of Doublet's Latin style, I can only emphasize that the *impostors* method —which was carried out as last above on p. 228— considerably decreases the likelihood of the charter's composition by any other author than Suger, unless Doublet was a very gifted forger indeed. The benefit of statistics may well be clear in this scenario. It allows to create a distance between the reader and Brown's arguments by allotting a likelihood to her hypothesis, whereby we are forced to reassess whether or not a specific case was actually approached from an apt theoretical angle in the first place.

What Van de Kieft and Brown have in common is that the mechanisms behind their attribution are inextricably linked to a close reading interpretation, their convictions on the text's intention and —indeed— a prefigured framework from which they approach twelfth-century authorship. Both scholars reflect on the problem from a cultural viewpoint that is 'author-centred,' in which medieval authors are held accountable for the truth- and falsehoods they proclaim. But there are also differences between them. Van de Kieft's implication is that the charter was given credence to by contemporary twelfth-century readers and manufactured with the purpose of deception. This is an idea which scholars such as Brown and Hohler cannot accept, out of the conviction that also for medieval readers the charter was "clearly an audacious and unmitigated hoax, inspired by even greater flights of fancy than the [*Pseudo-Turpin*] and focused far more centrally than that narrative on France's patron saint."[92] This may conjure up parallels in the twelfth century where authors have been thought to be 'poking fun' at the recurrent myths of their time. One may think, for instance, of the well-known *Institutio Traiani* supposedly written by Plutarch to the emperor Trajan, a *speculum* for an emperor, which John of Salisbury incorporated in books 5 and 6 of his *Polycraticus*. Scholars nowadays agree that the actual author was none author than John himself.[93] In discussing the forgery, Harold Love remarked that "John may have assumed that the *cognoscenti* among his audience would have picked up the

in the Middle Ages should not, in the absence of a true sense of literary individuality, be termed plagiarism [...] The force of Constable's argument is considerably blunted by his convincing demonstration that in the twelfth and thirteenth centuries such practice was viewed with disfavor," see Brown, "*Falsitas pia sive reprehensibilis*," 103, n. 7. The piece which Brown reacts to here is Constable, "Forgery and Plagiarism."

[90] "Aiming to defend the reputations of the medieval church and its clerics, the apologists of medieval forgers have taken two tacks, concentrated first on the products and second on the producers," see Brown, "*Falsitas pia sive reprehensibilis*," 103.

[91] See p. 218 above, and Brown, "Saint-Denis and the Turpin Legend."

[92] Brown, "Saint-Denis and the Turpin Legend," 54; or also see 71: "If the forger worked in the twelfth or thirteenth century, he had a keen sense of humor and took pleasure in poking fun at the pretensions of the house and its abbots, going so far as to assert that no kings should be crowned save at the abbey." For Hohler's take on the matter see the comment I have quoted on p. 212, which was taken from Hohler, "A Note on *Jacobus*," 37.

[93] "In Books Five and Six, John purports to exploit a manual of political theory composed by Plutarch for the Roman

fraud and been amused by it." He continues this train of thought in stating that "his readers would in any case have been aware that in the world around them much more self-interested acts of forgery were a common occurrence, particularly those of faked deeds, faked charters and faked wills."[94] Is the donation of Charlemagne, then, an intellectual game? Or should it be categorized under this latter group of "self-interested acts of forgery," repeating Love's terms above? Suger did not write for a similar audience as John, nor am I convinced that he was the man for 'jokes' of the sort mentioned here.

Derek Pearsall, in a contribution titled "Forging Truth," reflected specifically on fabricated charters of this kind.[95] Pearsall emphasized that whether or not medieval readers would have recognized the contents of a charter such as the donation as fictional is irrelevant. More on the mark is if that recognition came with the same sort of 'shock' as it did for contemporary historians and literary scholars. Details within the document that might have increased or decreased its 'authenticity' were a trifle matter, since the document was already true —so to speak— before it was 'formally' written. Even though Charlemagne's legendary stories were generally enacted differently than legislative charters, they had a function within society that could manifest itself in many different ways and on many levels. Such seems to be suggested, for instance, by the annual fair of Lendit commemorating Charlemagne's translation of the relics (already mentioned on p. 208), or by the tradition of the four bezants.[96] The charter plays with the language of a game that was already well understood, indeed, accessing and employing a kind of code woven within the fabric of society. In this light, Pearsall moreover argued that "a forged charter is indeed sometimes no more than a repetition, from memory or tradition, of a genuine document lost or destroyed."[97]

We have no historical evidence if the donation of Charlemagne was ever proffered for consideration and, if so, when and where. The charter's demands were clearly never legally reinforced, nor was it granted the royal acceptance it appears to aim for.[98] The additional bit of stylometric evidence brought forward in this chapter, where a 'ridiculous' charter's style can be linked to an authority such as Suger, places many of the old questions firmly on the agenda again. When considering that its authorship is momentarily the only hold we have on this slippery text, we inevitably have to

emperor Trajan. He cites this source explicitly (5.1) as 'a letter' (*epistola*) and later refers to it as 'a little book' (*libellus*) whose title is *Institutio Traiani*. However, John's definitive *extat*, 'there exists,' has not only been called into question, but discredited outright and called 'a fiction' devised by our author to conceal the personal character of his political views, just as he employed pseudonyms so widely in satirical passages. Although a case has been made for the existence of such a lost work by Plutarch, prevailing scholarly opinion denies the authenticity of the *Institutio Traiani* and accuses John of Salisbury of forgery," in Ronald E. Pepin, "John of Salisbury as a Writer," in *A Companion to John of Salisbury*, ed. Christophe Grellard and Frédérique Lachaud, Brill's Companions to the Christian Tradition 57 (Leiden: Brill, 2015), 176.

[94] Love, *Attributing Authorship*, 179.

[95] Pearsall, "Forging Truth."

[96] Levillain demonstrated that the fair's origins are early twelfth-century. See Levillain, "Essai sur les origines du Lendit."

[97] Pearsall, "Forging Truth," 5.

[98] As emphasized in Barroux, "L'Abbé Suger," 24.

ask ourselves what were Suger's motivations in order to interpret this text's original function and context. The most likely answer seems to be that Suger attempted to forge ties with the Capetian dynasty, but whether or not he received the response he hoped for remains open to question.

# 8

# Love to the Letter: Heloise and Abelard

## 8.1 Two Letter Exchanges

Central in this chapter are the two well-known letter exchanges of debated authorship in which Heloise of Argenteuil († 1164) and/or Peter Abelard (1079– 1142) are involved. The first is their conventionally 'accepted' letter exchange, consisting of eight letters, which sets out with Abelard's *Historia Calamitatum* (Ep. 1), and culminates with the institutional *Regula* for the Paraclete (Ep. 8), the monastery founded by Abelard in 1122 and seven years later presided over by Heloise as prioress.[1] The second is the *Epistolae duorum amantium* (hereafter *EDA*),[2] alternatively encountered as the 'lost love letters of Heloise and Abelard' ever since Constant Mews popularized the theory that Heloise and Abelard were the text's original authors.[3] For the sake of clarity, throughout the next pages, we will distinguish the first collection of letters from the second by consistently identifying them as respectively 'Heloise and

[1] As edited last in Petrus Abaelardus and Heloisa Argentoliensis, *The Letter Collection of Peter Abelard and Heloise*, ed. David E. Luscombe, trans. David E. Luscombe and Betty Radice, Oxford Medieval Texts (Oxford, UK: Clarendon Press, 2013). Before Luscombe's edition, especially Muckle's (C.S.B.) editions of the letters as published in *Mediaeval Studies* (1950–5) were considered the standard. Note that here we adopt the numbering of the letters as maintained in Luscombe's edition (and that of François d'Amboise and André Duchesne's earliest edition of 1616, the *Opera Omnia* reprinted in *PL* 178:113-314), instead of that maintained in Muckle's, where the numbering of the letters only started with Heloise's first response letter following up on Abelard's *Historia Calamitatum*. Here, the *Historia Calamitatum* equals Ep. 1 of the letter exchange. For Muckle's edition, see Joseph T. Muckle, "Abelard's Letter of Consolation to a Friend (*Historia Calamitatum*)," *Mediaeval Studies* 12 (1950): 163–213; "The Personal Letters Between Abelard and Heloise," *Mediaeval Studies* 15 (1953): 47–94; and "The Letter of Heloise on Religious Life and Abelard's First Reply," *Mediaeval Studies* 17 (1955): 240–281.

[2] See Könsgen, *EDA*.

[3] See Constant J. Mews, *The Lost Love Letters of Heloise and Abelard. Perceptions of Dialogue in Twelfth-Century France*, 2nd ed., trans. Neville Chiavaroli and Constant J. Mews, The New Middle Ages (New York: Palgrave, 2008 (1999)); the theory of Heloise and Abelard's authorship had already been formulated as early as Könsgen's edition, as suggested by the edition's second title (see n. 2 above). The English translation and Mews' English publication of the same hypothesis could count on much more attention than its German predecessor. This is emphasized in Ziolkowski, "Lost and Not Yet Found," 177.

Abelard's letter collection' (or 'Heloise-Abelard letter collection') and the *EDA*. We come back to the *EDA* later onward, and in the first paragraphs focus first on the letter collection, which —despite all the potential pitfalls—, remains a unique source for gaining an acquaintance with Heloise and Abelard.

## 8.2 The Letter Collection

Not in the least due to their romanticized reception from medieval times up until today, there are few figures more emblematic for twelfth-century history and Latinity than Heloise and Abelard. Abelard, the intellectual Parisian master with a rich oeuvre containing letter writing, poetry, theology, logic, and ethics, has taken on "iconic" proportions,[4] and especially after the recent surge of feminist literary criticism, Heloise has become the "model of the French public female intellectual, the first and one of a very few such figures before modern times."[5] The pair's letter corpus almost in every way forms witness of the most fascinating aspects of the twelfth century, such as the rise of the schools, the proliferation of monastic ideals (and the persecution of some of them), the tensions between state and church, the adherence to literary and stylistic models of epistolary composition, the rise of fictionality, the discovery of the individual, and —indeed— a rare glimpse of the interaction between men and women of letters at the time. Simultaneously, a paradox emerges. No medieval work stands on par with Heloise and Abelard's letters when it comes to universal appeal and familiarity. And yet, at the same time, no other medieval text is as divisive and remains as far out of reach.

Here is a text with no contemporary manuscript, without an absolute date of composition or composition context. This text has no clear history of circulation or transmission. We lack knowledge on its intended audience, it adheres to no uniform genre or tone, and describing it as either 'factual,' 'fictional' or even 'deliberately misleading' just does not feel right. And yet, this particularly entertaining and valuable piece of Latin literature is —aside from a few exceptions— our best hope of harvesting historical data about the lives of Heloise and Abelard and their literary undertakings. It is a story which either they themselves (or one of them?) meticulously crafted, and *wanted* us to read, or —if we believe a minority of scholars— it is a story crafted by a forger, whose intentions we cannot even begin to imagine.[6] Before turning to these long-standing authenticity debates below, it suffices to say that regardless which of many authorship hypotheses one believes more credible, no serious scholar suggests that the letter corpus is an objective and historically representative account of

[4] Verbaal, "Trapping the Future," 187–8.

[5] Taken from the introduction to Bonnie Wheeler, ed., *Listening To Heloise. The Voice of a Twelfth-Century Woman*, The New Middle Ages (New York: Palgrave, 2000), xvii.

[6] These scholars and their respective theories of why the Heloise-Abelard collection is a forgery are discussed further down on p. 245.

the lovers' lives.

### 8.2.1 *Historia Calamitatum*

Most of Heloise and Abelard's story is recounted in the *Historia Calamitatum*, a lengthy (auto?)biographical account narrating the events of Abelard's life at the opening of the letter exchange, and supposedly written by Abelard himself c. 1132 with the purpose of consoling his male friend.[7] Abelard begins his life story with his education in Paris c. 1100, the founding of his first schools in Melun and Corbeil soon thereafter, and his slow rise to fame as a dialectician and his education in Laon.[8] It is at the peak of his fame as a teacher in Paris c. 1116/7 that he is asked by Heloise's uncle Fulbert, a canon of Notre-Dame, to become his niece's tutor, upon which the pair of lovers begin their affair. In Abelard's own words, Heloise is not only conspicuous in good looks, but even more supreme (and therefore more attractive to him) due to her abundant knowledge of letters, a rare feat for the female sex (*in mulieribus est rarius*).[9] As the two lovers' infatuation becomes more intense, Abelard even succeeds at arranging accommodation for himself under Fulbert's roof and close to Heloise, under the pretense that having to run his own household distracts him from study. Fulbert, who is portrayed as stingy, avaricious and possessively fond of his niece, cannot believe his own luck when the most renowned philosopher of Paris knocks at his door to ask for lodging.[10]

However, when he catches wind of an affair between the two, he tries to separate them, but to no avail. Tensions rise high when Heloise becomes pregnant, upon which Abelard temporarily sends her to his birth region (Le Pallet in the region of Nantes, Brittany) to live with his sister Denise until she gives birth to their son Astralabe (still in 1116).[11] In order to soothe Fulbert, Abelard guarantees his honest intentions, and proposes a compromise, namely a hidden marriage, in which Abelard's proper reputation as *magister* is safeguarded and in which Heloise is not left to herself but provided for under his care.[12] Against Heloise's will,[13] the marriage takes place, but Fulbert fails to meet Abelard's demands to keep the affair a secret.[14] In an attempt to protect Abelard's reputation, Heloise consistently contradicts the truthfulness of her uncle's assertions in public, which causes the latter to fly into rage against her. Having witnessed Fulbert's short-fused reactions, Abelard decides to protect Heloise by providing her shelter in the convent of Argenteuil, where she had formerly been

[7] The identity of this friend remains unknown.

[8] Petrus Abaelardus and Heloisa Argentoliensis, *Letter Collection*, §1–12, 2–21.

[9] Ibid., §16, 24.

[10] Ibid., §16–7, 24–9.

[11] Ibid., §21, 30–3.

[12] Ibid., §23, 32–5.

[13] Ibid., §24–6, 34–43.

[14] Ibid., §28, 42–5.

raised and educated as a child. Fulbert interprets Abelard's intercession on Heloise's behalf as treasonous. This sparks his well-known vengeful reaction, in which he sends a band of henchmen to brutally castrate Abelard in the middle of the night.[15]

The tragic event separates the lovers (c. 1118/9). Abelard takes the habit in the monastery of Saint-Denis under abbot Adam, and Heloise takes the veil in the convent of Argenteuil.[16] In the ensuing parts of the *Historia Calamitatum*, Heloise is temporarily left out of the picture, as Abelard recounts of the further humiliations and persecutions by his rivals.[17] Abelard's troubles eventually lead him to Troyes, where he builds a simple oratory in devotion of the Trinity on a bit of land, soon to be known as the Paraclete. Students quickly flock to Abelard's settlement, taking on the simple life style of the ancient philosophers.[18] However, jealous conspirers who witness Abelard's success again compel him to flee his settlement and seek refuge in Saint Gildas-de-Rhuys, where he is unanimously elected abbot, although the barbarous ways of the monks there displease him.[19] In April 1129, Suger of Saint-Denis gains possession of Argenteuil, where Heloise resided, expelling the nuns and replacing them with monks from his own abbey. Abelard sees the opportunity to reinstall the Paraclete with Heloise as its prioress.[20] This provides him with a convenient excuse to visit her more frequently, but once again, Abelard's sporadic visits to the nuns of the Paraclete attract the attention of his rivals, wherefore Abelard is forced to go on the run once more. The *Historia Calamitatum* ominously ends with Abelard as a fugitive, feeling threatened by his opponents at every turn.[21]

The details of the couple's biographies become less clear after the events narrated in the *Historia Calamitatum* have taken place. Most information on the ensuing years, c. 1133–8, can be derived from Epp. 2–8 (allegedly written around that time), but as Heloise and Abelard enter into a conversational mode, the *Historia Calamitatum*'s first-person, chronologically structured narrative is abandoned for a correspondence which is more anecdotal and gives fewer clues for a reconstruction of the historical context. In these years Abelard took up teaching again in Paris at the Montagne Sainte-Geneviève, since in 1136, "the year after the illustrious king of the English, Henry, the Lion of Justice, departed his life," he certainly taught John of Salisbury.[22]

[15] Petrus Abaelardus and Heloisa Argentoliensis, *Letter Collection*, §29, 44–5.

[16] Ibid., §31, 42–5.

[17] The first is by the envious duo Albericus of Rheims and Lotulf of Lombardy, the followers of the theologians William of Champeaux and Anselm of Laon whom Abelard had insulted. The two accuse Abelard of heretic reasoning in his treatise on the Trinity, after which Abelard is condemned and ordered at an enquiry in Soissons in 1121 to burn the book in public. After this, in 1122, Abelard provokes further scandal as the monks of Saint-Denis react indignantly to his refusal to accept the identity of the patron saint Dionysius the Areopagite, which he had based on his reading of Bede's *Expositio Actuum Apostolorum*. See ibid., §32–51, 50–81.

[18] Ibid., §52–9, 80–93.

[19] Ibid., §60–2, 92–9.

[20] Ibid., §63–4, 98–103. On the expulsion from Argenteuil, and Suger's role, see the previous chapter on p. 214.

[21] Ibid., §65–75, 102–21.

[22] "Cum primum adulescens admodum studiorum causa migrassem in Gallias, anno altero postquam illustris rex An-

On occasions, he visited Heloise in the Paraclete, or corresponded with her.

Those scholars who support the theory of Heloise and Abelard's literary collaboration on the letters —to which we will return below— believe that in this time span the corpus was gradually composed and completed by 1136/7. The years after, which are the last of Abelard's life († 1142), probably had the worst of all his 'calamities' in store. In 1141, Abelard is condemned at the council of Sens,[23] for heresies first signalized by William of Saint-Thierry in 1140 in his *Disputatio Adversus Petrum Abaelardum*[24] and sent with an accompanying letter to Bernard of Clairvaux and Geoffrey of Chartres († 1149). Consequentially, Abelard is forbidden by papal order to write on theological matters, and when forced to retreat, is accepted by a compassionate Peter the Venerable at Cluny. Peter sends Abelard to the Cluniac priory of Saint-Marcel-sur-Saône, where the latter dies (apparently from a disease) in 1142.

## 8.2.2 The Authenticity Debate

### Material Evidence

As mentioned already in passing, the most bothersome lack of evidence for the Heloise-Abelard collection has been the absence of a twelfth-century manuscript. The earliest copy of the text was drawn up in the early thirteenth century shortly before 1236/8, in a codex known as MS Troyes 802. Unfortunately, this means that our best manuscript stems from a period some hundred years posterior to the original completion date around 1136/7. Or, at least, that is the assumption if Heloise and Abelard's text is indeed the twelfth-century text it purports to be, and not a thirteenth-century forgery.

Since this unique manuscript is of considerable importance in the authenticity debate, a better understanding of its material structure is necessary.[25] Troyes MS 802 is a composite codex, consisting out of three codicological units. Units two and three,

glorum Henricus leo iustitiae rebus excessit humanis, contuli me ad Peripateticum Palatinum, qui tunc in monte sanctae Genouefae clarus doctor, et admirabilis omnibus praesidebat. Ibi ad pedes eius prima artis huius rudimenta accepi, et pro modulo ingenioli mei quicquid excidebat ab ore eius tota mentis auiditate excipiebam ," see Joannes Saresberiensis, *Metalogicon*, 2.10; "When first as quite a young man I went abroad to study in France —it was the year after the illustrious king of the English, Henry, the Lion of Justice, departed his life— I betook myself to the Peripatetic of Le Pallet, who at that time presided at Mont Sainte-Geneviève, a famous teacher and admired by all. There, at his feet, I received the first rudiments of this art, and, to the limited extent of my poor intellect, with all eagerness of mind snatched up every crumb that fell from his lips." Translation taken from Joannes Saresberiensis, *Metalogicon*, trans. John Barrie Hall and K.S.B. Keats-Rohan, *CC CM* in translation 12 (Turnhout: Brepols, 2013), 2.10, 198. For John's education at the schools, see Cédric Giraud and Constant Mews, "John of Salisbury and the Schools of the 12th Century," in *A Companion to John of Salisbury*, ed. Christophe Grellard and Frédérique Lachaud, Brill's Companions to the Christian Tradition 57 (Leiden and Boston: Brill, 2015), 31–62.

[23] The council of Sens has for some time been falsely dated on 2 June 1140, but in the meantime consensus has arisen that the council took place in 1141. For extensive argumentation on the dating and the underlying political dimensions to the council, see Mews, "The Council of Sens (1141)"; and Verbaal, "Council of Sens."

[24] Guillelmus de Sancto Theodorico, *Disputatio adversus Petrum Abaelardum*, in *Guillelmi a Sancto Theodorico opera omnia*, ed. Paul Verdeyen, vol. 5, *CC CM* 89A (Turnhout: Brepols, 2007), 13–59.

[25] For most of what follows, David Luscombe's introduction to the recent 2013 edition is authoritative, referenced in n. 1. The manuscript can be consulted online at https://portail.mediatheque.grand-troyes.fr/

of no further interest here, are Carolingian.[26] The first unit, however, is thirteenth-century,[27] and comprises Heloise and Abelard's letter collection (Epp. 1–8, fos. 1$^{ra}$–88$^{vb}$). As can be inspected in a schematic visualization of the manuscript in fig. 8.1, this first unit extends further than the collection (fos. 89$^{ra}$–103) and is complemented by five additional segments that have for a long time been overlooked by scholars.

These segments consist of legislative and canonistic texts, drawn from disparate sources. Their common theme is the regulation of female religious life through a series of statutes, instructions and directives.[28] For some time, they were known under the misleading title *Excerpta ex regulis Paraclitensis monasterii*, which d'Amboise and Duchesne had installed in their *editio princeps* of 1616.[29] Only in 1962, due to an important article by Damien Van den Eynde,[30] and due to renewed critical investigation into the collection's authenticity after 1972,[31] have they attracted the attention they deserve. Not in the least because they can be dated rather precisely (see fig. 8.1), therefore contributing in the heated debates around the principal manuscript's history and the letter collection's original composition date.[32] The *Excerpta* have also been taken to illustrate the agenda underlying the letters. A modern audience is often distracted by the appeal of Heloise and Abelard's love affair in the *Historia Calamitatum*

[26] They contain the Frankish chronicle of Fredegar (fos. 104–55) and John Scotus Eriugena's (815–c. 877) translations of works by pseudo-Dionysius (fos. 156–241).

[27] Dalarun confirmed the codicological unity of these texts by referring to the "quality of the parchment, the relative constancy of the gatherings' assembly, the presence of four signatures on the gatherings which are still visible, the homogeneity of the ruling schema, the fact that two scribes of the same century take turns in the second column of fol. 55$^{v}$ in the middle of the seventh letter, the fact that the thirteen texts written on 103 leafs succeed each other without the transitions between them ever corresponding to a change of gathering, and also the interventions of a single artist for the watermark initials, and of a same reviser for the whole."My translation from Jacques Dalarun, "Nouveaux aperçus sur Abélard, Héloïse et le Paraclet," *Francia* 32, no. 1 (2005): 21.

[28] For a meticulous and well-referenced state of the art concerning the manuscript's composite character and the implications, illustrated with facsimiles, see the first pages of ibid., especially 19–31.

[29] Reprinted in *PL* 178:313–25. D'Amboise and Duchesne based their edition on three exemplars, respectively from Nantes, Saint-Victor (Paris) and the Paraclete. Only the latter contained the *Excerpta*. Also see Damien P. Van den Eynde, "En marge des écrits d'Abélard: les *Excerpta ex regulis Paracletensis monasterii*," *Analecta Praemonstratensia* 38 (1962): 70; it might be the case that the Paraclete exemplar used in d'Amboise and Duchesne's 1616 edition was in fact none other than Troyes MS 802. This is an unresolved debate, summarized in Luscombe's introduction to Petrus Abaelardus and Heloisa Argentoliensis, *Letter Collection*, cxxiii–cxxiv.

[30] Van den Eynde, "En marge des écrits."

[31] John F. Benton, "Fraud, Fiction and Borrowing in the Correspondence of Abelard and Heloise," in *Pierre Abelard—Pierre le Vénérable: les courants philosophiques, littéraires et artistiques en Occident au milieu du XIIe siècle*, Colloques internationaux du Centre national de la recherche scientifique 546 (Paris: Éditions du Centre National de la Recherche Scientifique, 1975), 467–512.

[32] So, for instance, the currently dominant *termini post* and *ante quem* of the manuscript —after 1230 and before 1236/8— have been conjectured by aid of these neighbouring documents. Troyes MS 802 was drawn up after the council of Rouen in 1231, for it integrates two of its canons, and even provides a superior reading. For a side by side comparison between the Rouen canons in Troyes MS 802 and those maintained in MS Avranches, Bibliothèque municipale, 149, see John F. Benton, "The Paraclete and the Council of Rouen of 1231," *Bulletin of Medieval Canon Law* 4 (1974): 36–7; for the *terminus ante quem*, Van den Eynde's dating of the Premonstratensian statutes before 1236/8 is significant, see Van den Eynde, "En marge des écrits," n. 37, 83. MS Troyes 802 provides an earlier and anterior redaction of the most recent of the statutes in MS Avranches 149, which had been completed by 1236/8. For a helpful summary (with references) of the argumentation underlying this dating, see Dalarun, "Nouveaux aperçus," 30–1.


| | folios | titles | date |
|---|---|---|---|
| 1 | fos. 1ra–88vb | *Letter Collection* (Epp. 1–8). | — |
| 2 | fos. 89ra–90va | *Institutiones nostrae*. | c. 1142/47 |
| 3 | fos. 90va–93ra | Ivo of Chartres's *Panormia* 3.187–215 (*PL* 161:1175–81). | c. 1095 |
| 4 | fos. 93ra–93rb | Three canons, two of which from a council of Rouen (1231), that have a bearing on female religious life. | c. 1231 |
| 5 | fos. 93rb–94rb | Eleven Premonstratensian statutes. | 1174–1236/38 |
| 6 | fos. 94va–102va | *Regula sanctimonialium* of the diet of Aachen, 7–28. | 816 |

Figure 8.1: Schematic visualization of the configuration of MS Troyes 802 (fos. 1–103), the oldest and most reliable manuscript containing Heloise and Abelard's letter collection (Epp. 1–8).

and ensuing letters. By contrast, a medieval audience would have been more sensitive to the document's climactic build-up toward the stipulations of monastic life at the end. That extensions in MS Troyes 802 appear there may serve to emphasize these passages' interest to later generations of readers.

There is one particularly interesting excerpt which deserves some elaboration: the *Institutiones nostrae*, or the *Paraclete Statutes*, have been argued to be ascribable to Heloise. This short customary, formulating the observances of an unspecified female convent, was written some time during the 1140s. It has been taken to correspond to a rule for the Paraclete,[33] written at a time when Heloise was the monastery's prioress. This ascription to Heloise has to my knowledge not been contested, neither has it been conclusively proven. Chrysogonus Waddell, the text's most recent editor, defended the position of Heloise's authorship in 1987, with much evidence.[34] As we will come to discuss below, the statutes' provenance from the Paraclete and their possible attri-

[33] This has been the case as early as d'Amboise and Duchesne's edition of 1616. See Julia Barrow, Charles S.F. Burnett, and David E. Luscombe, "A Checklist of the Manuscripts Containing the Writings of Peter Abelard and Other Works Closely Associated with Abelard and his School," *Revue d'histoire des textes* 14–5 (1984–5): 284–5; Van den Eynde maintained that the *Institutiones nostrae* pertained to the Paraclete, due to the invocations to the Holy Spirit, and the way in which it borrows heavily from Abelard's Rule in Ep. 8. See Van den Eynde, "En marge des écrits," especially on 72–3. Waddell showed that the *Institutiones* relied on Cistercian usages written before 1147. See Chrysogonus Waddell, ed., *The Paraclete Statutes: Institutiones Nostrae: Troyes, Bibliothèque Municipale, MS 802, ff. 89r–90v*, Cistercian Liturgy Series 20 (Trappist, Kentucky: Cistercian Publications, 1987), in his commentaries on 28-65, 87, 92–102 and 199–203.

[34] "Given the probability that the text really does date from early in the Paraclete's history, it would seem that only the abbess of the Paraclete would have had the authority to draw up an official statement of the Paraclete's characteristic observances. Certainly, she had the competence to do so, and hardly stood in need of outside help to formulate the document," see Waddell, *The Paraclete Statutes*, 202. It should be noted, on the other hand, that Waddell was not won for the idea that Heloise wrote Epp. 2, 4 and 6 in the correspondence. See ibid., 41: "The entire correspondence, Letters I–VIII (with the Rule) is from the pen of a single writer; and this writer is Peter Abelard."

bution to Heloise is no triviality. The *Institutiones nostrae*'s divergence from some of the stipulations made by Abelard in his own rule for the Paraclete (Ep. 8) has given some scholars an argument to plead in favour of the collection's inauthenticity.[35]

Be that as it may, not Troyes MS 802 or any other available witness of the letter collection conclusively proves or disproves that the letter collection is what it purports to be: a correspondence written by Heloise and Abelard in the course of the 1130s. If the letter corpus is authentic, the lack of a contemporary manuscript is a frustrating coincidence. Very few of Heloise and Abelard's other texts lack a twelfth-century copy.[36] Those scholars who reject the corpus's authenticity (see below), have often invoked this conspicuous lack of documentary evidence and odd physical detachment from the twelfth century as an argument in favour of forgery.

Not only do we lack a physical manuscript, we cannot even determine if such a manuscript ever existed. It is unclear whether or not the letters were actually 'published,' which, in the medieval sense, would mean as much as the autographs' contemporaneous transcription and/or circulation outside of a private collection or archive. Although there is sufficient evidence to prove that contemporaries were familiar with Heloise and Abelard's story, such as the letter of Fulco (Fulk) of Deuil, or that of Abelard's former master and adversary Roscelin of Compiègne (c. 1050–c. 1124),[37] there is no evidence to show that contemporaries had read the Latin letters that lie before us today. More importantly, there is no evidence on who collected them and when, and who had the last hand in them. Indeed, our earliest proof of a 'material' existence of a letter corpus comprising the correspondence of Heloise and Abelard dates —again— from the thirteenth century, when the French poet Jean de Meun (c. 1240–c. 1305) discovered a copy and translated it around 1280. Not surprisingly, there have been supporters of the forgery theory who drew their own conclusions from the collection's first 'tangible' appearance in the hands of the gifted writer Jean de Meun.[38]

## Text-Internal Criticism: History and Fiction

As material evidence is likely to remain scarce, scholars have as early as the 1800s[39] fallen back on their proper convictions of how the letter corpus is best interpreted,

[35] Benton, "Fraud, Fiction and Borrowing," see p. 244 in this chapter, where Benton's paper is discussed.

[36] Joseph T. Muckle, "The Personal Letters Between Abelard and Heloise," *Mediaeval Studies* 15 (1953): 49.

[37] The letter, titled as "Epistola ad Abaelardum"and written c. 1119/20, is printed in *PL* 178:371–6, and critically edited in Roscelinus Compendiensis, "Der Brief Roscelins an Abälard", in *Der Nominalismus in der Frühscholastik*, ed. Josef Reiners, *Beiträge zur Geschichte der Philosophie des Mittelalters* 8.5 (Münster: Aschendorffschen Buchhandlung, 1910), 63–80.

[38] Silvestre maintained the opinion that the whole dossier was forged by Jean de Meun. See Hubert Silvestre, "Die Liebesgeschichte zwischen Abaelard und Heloise: der Anteil des Romans," in *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, vol. 5, Monumenta Germaniae Historica Schriften 33 (Hannover: Hahnsche Buchhandlung, 1988), 121–65.

[39] Deborah A. Fraioli, "George Moore and Scott Moncrieff: An Unknown Chapter in the the Authenticity Debate of the Letters of Abelard and Heloise," *Forum for Modern Language Studies* 54, no. 2 (2017): 176.

dated and contextualized. But to dissect the correspondence in an attempt to segregate what is authentic from what is false has proven challenging, if not in some cases disastrous, to say the least. It is important to consider how already in Heloise and Abelard's lifetime, legendary versions of the lovers' story circulated.[40] This has always placed the historical reliability of the events narrated in the letter collection under high tension, and equally so the reliability of its narrators. Numerous scholars have shown how the text indulges in celebrating its own textuality. So Abelard does not merely speak as Abelard, but imitates normative author(itie)s such as Ovid, Augustine or Jerome. The same applies to Heloise (or is it Abelard?), who lengthily cites from Jerome's *Contra Jovinianum*[41] and from Seneca's letters.[42] Heloise speaks as Vergil's Dido in the *Aeneid*, as an Ovidian heroine from the *Heroides*,[43] and as Lucan's Cornelia in the *Bellum Civile*.[44] The love story of Heloise and Abelard does not work within a textual vacuum and cannot be termed an "intimate and personal" correspondence.[45] It is a literary text, with a function toward an audience, a function which the defenders of its authenticity have often found in legitimizing the Paraclete by means of a foundational record.

Indeed, the letter corpus goes to great lengths in identifying itself as a story in which the world's involvement is essential. On pivotal occasions *mundus* becomes a present onlooker and participant to the *Historia Calamitatum*. When Abelard arranges their marriage, for example, Heloise's ensuing premonition that a tragic event is at hand demands affirmation of the world as audience (*sicut universus agnovit mundus*).[46] When that tragic event occurs, which is Abelard's emasculation, it is described as a world-shocking event (*quam summa ammiratione mundus excepit*).[47] This text needs a readership to incite its central themes such as public disgrace, envy and betrayal. Abelard literally sees the entire world conspiring against him. But at the same time, in between all the contempt for and by the world, he also needs the world's *ammiratio* in order for his story to be operational. The world, accused for Abelard's misfortunes, is the *Historia*'s adversary.

[40] Long before later well-known romanticizations by Alexander Pope (in 1717) and Jean-Jacques Rousseau (in 1761). See, for instance, Luscombe's introduction to Petrus Abaelardus and Heloisa Argentoliensis, *Letter Collection*, xvii; and Peter Dronke, *Abelard and Heloise in Medieval Testimonies*, W. P. Ker Memorial Lecture 26 (Glasgow: University of Glasgow Press, 1976).

[41] Deborah A. Fraioli, "The Importance of Satire in Jerome's *Adversus Jovinianum* as an Argument against the Authenticity of the *Historia Calamitatum*," in *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, vol. 5, Monumenta Germaniae Historica Schriften 33 (Hannover: Hahnsche Buchhandlung, 1988), 167–200.

[42] From the letters to Lucilius.

[43] Phyllis R. Brown and John C. Peiffer II, "Heloise, Dialectic, and the *Heroides*," in *Listening To Heloise. The Voice of a Twelfth-Century Woman*, ed. Bonnie Wheeler, The New Middle Ages (New York: Palgrave, 2000), 143–60.

[44] Peter von Moos, "Cornelia und Heloise," *Latomus* 34, no. 4 (1975): 1024–59.

[45] Muckle, when editing the texts in the 1950s, still fostered that impression. See Muckle, "The Personal Letters Between Abelard and Heloise," 47.

[46] Petrus Abaelardus and Heloisa Argentoliensis, *Letter Collection*, §27, 42.

[47] Ibid., §29, 44.

Most scholars will agree that this document cannot simply be read as a collection of reliable ego-documents, in which we hear the genuine aspirations of a twelfth-century couple. But from this impression a paradox emerges. The text's constructed nature and its purpose of being a foundational record seems to divulge a public function. On the other hand, evidence of a publication is lacking. This has been a most troublesome and unanswerable question. As has timely been emphasized, Heloise outlived Abelard by some twenty years. It is likely that she was the collection's possessor and, importantly, would have had most opportunity to assemble, revise and publish it.[48] Barbara Newman asserted that Heloise must have been embarrassed to do so, and might have feared the outburst of scandal at their disclosure.[49] Although this all remains plausible, one can sense this explanation is not entirely satisfactory. As Dronke has amply shown, the argument of 'moral condemnation' toward the lovers would derive from a misunderstanding of much more complex and ambiguous twelfth-century attitudes toward love and sex: none of Heloise and Abelard's contemporaries chided the pair for their affair.[50] Exactly what 'scandal' would Heloise have to fear from disclosing the collection, then? Was the letter collection not an urgent and necessary document to govern the community of the Paraclete and safeguard its customs?

The most intense renewal of the controversy, debating over the historical and fictional elements of the collection, was John Benton's "Fraud, Fiction and Borrowing" paper, delivered at an international conference held in the abbey of Cluny in 1972.[51] It put all of the above questions firmly on the agenda. Benton hypothesized that the letter corpus consisted only partially of authentic work by Abelard, but was for the remaining part a product of twelfth- and/or thirteenth-century forgers. As we came to mention already on p. 241, Benton thought the *Institutiones nostrae*'s discrepancies with Abelard's own *regula* in Ep. 8 to indicate that Abelard's Rule was a thirteenth-century forgery, meant to undermine the authority of an older Paraclete Rule, that is to say, the *Institutiones nostrae*. In that same year, Durant W. Robertson published a book in which the letter collection —including Heloise's letters— was presented as a piece of self-mockery written by Abelard alone.[52] Ever since Benton withdrew his argument in 1980, and was more inclined to designate Abelard as single author,[53] contemporary upholders of the forgery argument are outnumbered, but not wholly

[48] As emphasized in Étienne Gilson, *Héloïse et Abélard*, Études sur le moyen age et l'humanisme (Paris: Librairie philosophique J. Vrin, 1938), 34–5; Ludovic Lalanne, "Quelques doutes sur l'authenticité de la correspondance amoureuse d'Héloïse et d'Abailard," *La correspondence littéraire* 1 (1856): 30; and Dronke, *Women Writers*, 108.

[49] Newman, "Authority," 133–5.

[50] Dronke, *Medieval Testimonies*.

[51] Benton, "Fraud, Fiction and Borrowing."

[52] Durant Waite Robertson, *Abelard and Heloise* (London: Millington, 1972).

[53] John F. Benton, "A Reconsideration of the Authenticity of the Correspondence of Abelard and Heloise," in *Petrus Abaelardus (1079–1142). Person, Werk und Wirkung*, ed. Rudolf Thomas, Trierer Theologische Studien, 38 (Trier: Paulinus Verlag, 1980), 41–52.

refuted. Scholars such as Hubert Silvestre,[54] Deborah Fraioli[55] and Kees Schepers[56] still support the theory that the author(s) of the letter collection were some other author than Heloise and/or Abelard.

Even though none of the 'peculiarities' signalized by Silvestre, Fraioli and Schepers should be swept off the table, neither are they the hard historical evidence against inauthenticity as the authors make believe. Their reports on historical discrepancies or detected anachronisms rather illustrate that there is scholarly disagreement over what in fact constitutes 'normative' twelfth-century Latin text and/or behaviour. That there appears to be no consensus on this, and that the letter collection's authorship is determined not by objective parameters but by what may or may not fit one's individual bill of what twelfth-century literature looked like, provided a greater controversy to the academic community than the controversy over the letters' authorship.

## Ideology and Attribution

This awareness, which one will commonly find shared among most scholars studying Heloise and Abelard today, can be argued to be a fruitful consequence of the fierce *Gelehrtenstreit*[57] that has become an integral part of this text's history. As the Swiss scholar Peter von Moos argued in 1974, this state of affairs has gradually transformed the Heloise and Abelard debate to a meta-study, i.e. interest in the controversy around authorship rather than interest in authorship.[58] Countless hypothetical attributions circulate (or have circulated) for Heloise and Abelard, which are sometimes —for divergent reasons— 'silently' presented as accepted, when they are not. Needless to say that this has frustrated scholars, and only worsened a case for critically assessing these important historical figures and their literary works.

Perhaps the strongest undercurrent of scholarly ideology in the state of the art of Heloise and Abelard's letter collection has been misogynism. As was sharply exposed by Newman in 1992,[59] not Abelard, but Heloise suffered most from the state of uncertainty. She was denied the authorship over her letters numerous times,[60] reduced

[54] Silvestre, "Die Liebesgeschichte."

[55] Fraioli has taken Heloise's misunderstanding of this text —in which she argues "for free love rather than virginity and celibacy"— as an argument against the authenticity of the letter corpus, and a sign of the corpus's comic and satirical purpose. Fraioli argued "that the *Historia Calamitatum* and the accompanying letters are a literary forgery and an anti-Abelardian satire written by a third party," namely a contemporary adversary of the lovers. See Fraioli, "Importance of Satire"; Fraioli, "An Unknown Chapter."

[56] Kees Schepers, "Abelard's Exegesis of the Song of Songs in his Second Letter to Heloise," *The Journal of Medieval Latin* 27 (2017): 107–31.

[57] I use the term in reference to the subtitle in Peter von Moos, *Mittelalterforschung und Ideologiekritik. Der Gelehrtenstreit um Heloise* (München: Wilhelm Fink Verlag, 1974).

[58] von Moos, *Mittelalterforschung*; this point is also emphatically made in John Marenbon, "Authenticity Revisited," in *Listening To Heloise. The Voice of a Twelfth-Century Woman*, ed. Bonnie Wheeler, The New Middle Ages (New York: Palgrave, 2000), see 19: "Historians cannot take a purely objective, unideological standpoint, for there is none. But they can, argues von Moos, scrutinize their own and others' acknowledged or concealed ideologies."

[59] Newman, "Authority."

[60] Bernhard Schmeidler, "Der Briefwechsel zwischen Abälard und Heloise eine Fälschung?," *Archiv für Kulturgeschichte*

to Abelard's *fictio*, while he took credit for the correspondence as a whole. Divergent arguments were given for an explanation, but Newman, in wittily talking her readership through the state of the art and citing the inconsistencies, showed how none of them were based on evidence. Then again, the feminist reading of the collection and the advocacy for Heloise's voice cuts both ways. Newman's piece was a successful disclosure of the fallacies in scholars' attribution of the collection to Abelard, a perceptive dissection of a field caught in its own ideological entanglements. The question is if this has, in its own turn, not caused the debate to veer too passionately in the other direction. The rectification of Heloise's suppression became an aim in itself, where a contemporary, ideologically motivated greater good determined scholars' interpretation of the past.

## 8.3 The *Epistolae duorum amantium*

### 8.3.1 The *Lost Love Letters* (1999)

Before discussing further the role of ideology, Heloise's voice, and Heloise and Abelard's collaboration, we first need to discuss the second letter corpus of interest in this chapter. In his *Historia Calamitatum*, Abelard testifies that in the earliest days of their affair, he and Heloise secretively exchanged letters.[61] In 1974, two years after Benton's paper had caused disconcertion over the letter corpus' authenticity, Ewald Könsgen published an edition of an anonymous twelfth-century correspondence comprising 113 fragmentary love letters —partly in prose and partly in verse— between a famous teacher *V<ir>* and a gifted female student *M<ulier>*: the *Epistolae duorum amantium* (*EDA*).[62] Inevitably, the question coming to Könsgen's lips made it to the title of the edition: are these indeed *Briefe Abaelards und Heloises*? Strangely enough, one had to wait for Constant Mews's translation in 1999,[63] which did away with Könsgen's question mark and subscribed to the attribution of the *EDA* to Heloise and Abelard, before the hypothesis was taken seriously.[64] As pointed out by Newman, Könsgen's German edition, published in the middle of the controversy over the Heloise-Abelard collection, appeared to have missed its momentum.[65] Scholars were on their guard

11 (1914): 1–30; Charlotte Charrier, *Héloise dans l 'histoire et dans la légende*, Bibliothèque de la Revue de littérature comparée 102 (Paris: Champion, 1933); Robertson, *Abelard and Heloise*; Benton, "A Reconsideration"; Waddell, *The Paraclete Statutes*.

[61] "[...] Nosque etiam absentes scriptis internuntiis inuicem licere presentare, et pleraque audacius scribere quam colloqui," which translates to: "[...] Even when separated we could enjoy each other's presence by exchange of written messages in which we could write many things more boldly than we could say them." See Petrus Abaelardus and Heloisa Argentoliensis, *Letter Collection*, §16, Latin at 26, English translation at 27.

[62] Könsgen, *EDA*.

[63] Mews, *The Lost Love Letters*.

[64] Könsgen's edition generated far less interest than the English translation, as is emphasized in Ziolkowski, "Lost and Not Yet Found," 177.

[65] Barbara Newman, *Making Love in the Twelfth Century. Letters of Two Lovers in Context*, trans. Barbara Newman,

in making assumptions. Although the *EDA*'s discovery was picked up by some to revisit the authenticity of the Heloise-Abelard story,[66] it was only after the English publication of 1999 that the debate over this second letter exchange came in full sway.

When it comes to material evidence, matters are worse —far worse— for the *EDA* than for the Heloise-Abelard collection. Although scholars have generally consented to the correspondence itself being twelfth-century Latin (with exceptions below), its single witness is a fifteenth-century paper manuscript comprising an anthology or *summa dictaminis* copied in 1471.[67] The scribe responsible for the copying goes by name of Johannes de Vepria (c. 1445–c. 1518), who was a young Cistercian monk of twenty-five years old and later prior of Clairvaux (1480–99). The letters are considerably short and some of them are fragmentary. They may yet differ in length, from a couple of lines to more than a page. The text itself does not purport to be written by Heloise and Abelard —which makes this a fundamentally different authorship problem than that of the letter collection—, and for none of the letters can an explicit link to Heloise and/or Abelard be established. Also Johannes de Vepria, when copying the work, mentioned no name, nor attempted to make an attribution.[68]

Those who endorse an ascription to Heloise and Abelard, amongst whom Stephen Jaeger,[69] Sylvain Piron[70] and Newman,[71] find support for their claims by indicating a number of suggestive text-internal analogies between both couples' lives, and parallels in the lovers' phrasing, intertextual stock of resources, and preferred imagery. The pair of correspondents are a reputed *magister* and a young female student in an urban, intellectual milieu, probably the Parisian region, i.e. Île-de-France.[72] He is persecuted by jealous competitors, and she is a brilliant pupil. Striking as such overlaps may appear, even one of the staunchest proponents of the ascription, Jaeger, has recently

The Middle Ages Series (Philadelphia: University of Pennsylvania Press, 2016), xi.

[66] So, for instance, already two years after Könsgen's edition, Dronke referred to the *EDA* to demonstrate that scholars' refusal to accept that Heloise would have been capable of writing such sensual, profane letters as those under her name in the collection, apparently originates from an unfamiliarity of the Ovidian revival producing a tradition of amorous and erotic correspondences at the time. See Dronke, *Medieval Testimonies*, 24–5.

[67] Troyes, Bibliotheque municipale, MS 1452, fos. 159$^{r}$–167$^{v}$. For the contents of the preceding folios of the manuscript, see Könsgen, *EDA*, ix–xiv.

[68] Johannes de Vepria mentioned names and titles for all of the other works and extracts copied in MS Troyes 1452, but not for the *EDA*. See ibid., 76.

[69] C. Stephen Jaeger, *Ennobling Love: In Search of a Lost Sensibility*, The Middle Ages Series (Philadelphia: University of Pennsylvania Press, 1999), see 157ff. C. Stephen Jaeger, "*Epistolae duorum amantium* and the Ascription to Heloise and Abelard," in *Voices in Dialogue. Reading Women in the Middle Ages*, ed. Linda Steffel Olson and Kathryn Kerby-Fulton (Notre Dame, IN: University of Notre Dame Press, 2005), 125–66; and later C. Stephen Jaeger, "The *Epistolae Duorum Amantium*, Abelard, and Heloise: An Annotated Concordance," *The Journal of Medieval Latin* 24 (2014): 185–224.

[70] Piron is the French translator of the letters. See Sylvain Piron, ed., *Lettres des deux amants attribuées à Héloïse et Abélard* (Paris: Gallimard, 2005); and Sylvain Piron, "Heloise's Literary Self-Fashioning and the *Epistolae duorum amantium*," in *Strategies of Remembrance: From Pindar to Hölderlin*, ed. Lucie Doležalová (Newcastle upon Tyne: Cambridge Scholars Publishing, 2009), 103–62.

[71] See most recently Newman, *Making Love*.

[72] Könsgen, *EDA*, 91.

admitted that it constitutes 'soft' evidence.[73]

Skeptics, such as Peter Dronke,[74] Giles Constable,[75] Jan Ziolkowski,[76] and John Marenbon[77] have been bothered that Mews's attribution was, to say it in Marenbon's words, "a denial of the high scholarly standards he himself has set in twenty years of work on the area."[78] To their opinion, the conclusion of Mews and his associates was premature. There is a critical lack of material evidence and a considerable amount of scholarly disagreement on the *EDA* itself, which must make its potential relationship to the Heloise-Abelard a subquestion, not an answer. Indeed, the debate on the *EDA*'s function and whether or not they were composed by one or two authors is still ongoing. Although a majority of scholars contended that the styles of *M<ulier>* and *V<ir>* differ too extensively for single authorship to be sustainable, some still call it into question.[79] One may have qualms about accepting the *EDA* as forming a genuine correspondence. Peter von Moos, for instance, believed the exchange to have been composed by a single Italian author in the cultural climate of the *dolce stil novo* from the mid-thirteenth century.[80] Scholars have shown unease by the fact that Ovidian love poetry —drawing from the *Metamorphoses* and *Heroides*— knew a revival in the literary landscape of the late eleventh and twelfth century.[81] There are examples of lyrics and poems celebrating eroticized teacher-student relationships, both in urban and monastic contexts, and not all of them can be trusted to reflect reality.[82] The *EDA*'s affiliations to this tradition disallow an all too restrictive focus on Heloise and Abelard as candidate authors.[83] The exchange is moreover filled with dictaminal salutations, and indeed, as a whole it forms the closing piece of a manuscript containing models and formularies for epistolary writing. Are the passionate supporters of the *EDA*'s ascription not merely mistaking literary commonplaces for a genuine, realistic

[73] Jaeger, "An Annotated Concordance," 186.

[74] Dronke, *Medieval Testimonies*, from 24 onward.

[75] Giles Constable, "Sur l'attribution des *Epistolae duorum amantium*," *Comptes rendus des séances de l'Académie des Inscription et Belles-Lettres* 145, no. 4 (2001): 1676–93; and Giles Constable, "The Authorship of the *Epistolae Duorum Amantium*: A Reconsideration," in *Voices in Dialogue. Reading Women in the Middle Ages*, ed. Linda Steffel Olson and Kathryn Kerby-Fulton (Notre Dame, IN: University of Notre Dame Press, 2005), 167–78.

[76] Ziolkowski, "Lost and Not Yet Found."

[77] John Marenbon, "Lost Love Letters? A Controversy in Retrospect," *International Journal of the Classical Tradition* 15, no. 2 (2008): 267–80.

[78] Ibid., 280.

[79] Statistical analyses carried out by Giovanni Orlandi, Francesco Stella and Mike Kestemont appear to have settled the score for dual authorship. For Orlandi's analyses with $\chi^2$ (Pearson's chi-square), see Peter Dronke and Giovanni Orlandi, "New Works by Abelard and Heloise?," *Filologia mediolatina: Studies in Medieval Latin Texts and their Transmission* 12 (2005): 146–77. These results were reprinted, and augmented, in Ziolkowski, "Lost and Not Yet Found," 193–4. For Stella's conclusion, "Il principale è la conferma statistica che gli autori sono due, e che fra i due si registrano differenze non solo nelle preferenze di lessico 'automatico,' grammaticale, ma anche nella scelta di lessico astratto e affettivo," see Stella, "Analisi informatiche," 9. Jaeger referred to an unpublished workshop by Mike Kestemont and Kees Schepers, "Stylometric Exploration of the Implied Dual Authorship in the *Epistolae duorum amantium*," given at Louvain, April 2012. See Jaeger, "An Annotated Concordance," 188, n. 12.

[80] von Moos, "Die *Epistolae duorum amantium*."

[81] In Dronke, *Medieval Testimonies*; and Dronke and Orlandi, "New Works by Abelard and Heloise?"

[82] One can think of the Tegernsee collection and the Regensburg Songs (*Carmina Ratisponensia*) from Bavaria (Southern

correspondence?

### 8.3.2 The Authorship of Heloise

In Ziolkowski's eyes, far more was at stake for the supporters of the *EDA*'s attribution to Heloise and Abelard than simply settling an authorship dispute.[84] Mews's *Lost Love Letters of Heloise and Abelard* (1999) —not *Abelard and Heloise* (!)— had a wholly different agenda, namely "to redress the suppression and distortion Heloise had suffered in centuries of sexist scholarship."[85] After all, the authenticity of the 'secret' correspondence would validate that of the formerly discovered 'public' one, and its dual authorship by 'real' twelfth-century writers would assert that Heloise was more than a fictitious fantasy forged by Abelard. In the *EDA*, Mews found a preferable Heloise: a great author and lyricist, one that even outshone Abelard, since the <*M*>*ulier*'s style has often been more appreciated than that of the <*V*>*ir*. Has he overshot the mark? Has Newman, as well, in her support of the ascription for the sake of a greater good, —finding Heloise's voice—, fallen victim to what she had herself so vehemently criticized: ideology?[86] Von Moos's *Ideologiekritik* comes to mind —decades after he published it— as a work of great relevance.[87]

This brings us to the sensitive topic of Heloise's authorship. Very few of Heloise's writings may confidently be said to have remained immune from Abelard's influence. And yet, we have good reasons to maintain that she was an independent and gifted writer. So much becomes clear from the esteem granted to her by her contemporaries. Typically cited in this regard are the praises from the Augustinian canon Hugh Metel

Germany), see Dronke, *Medieval Testimonies*, 25; or of the literary tradition of Loire Valley with poets such as Marbod of Rennes (c. 1035–1123), Baudri of Bourgeuil (c. 1045–1130), Fulcoius of Beauvais († after 1083) and Hildebert of Lavardin (c. 1055–1133). Within this larger framework of the emergence of love poetry, Wim Verbaal argued that a twelfth-century fiction was most feasible, since the collection of the *EDA* "combines the genre of Marbod's love poems with the style of the fictive letter exchange we saw in Baudri," see Verbaal, "How the West was Won by Fiction," 199. Verbaal concludes by stating that "such was the story, then, of how truth became fiction, and this fiction, thanks to modern scholars, became fact again." Stella has provided a detailed account of the textual parallels between the *EDA* and —amongst other sources— authors such as Marbod, Fulcoius, Baudri and Hildebert, see Francesco Stella, "*Epistolae duorum amantium*: nuovi paralleli testuali per gli inserti poetici," *The Journal of Medieval Latin* 18 (2008): 374–97; the Tegernsee collection and part of the Regensburg Songs have recently been translated and commented upon in their relation with the *EDA*, see Newman, *Making Love*, 229–55 and 257–78.

[83] Especially Ziolkowski took issue at Mews's argument, which asserted that if there are no better known candidates than Heloise and Abelard, the text's attribution to them becomes reliable, see Ziolkowski, "Lost and Not Yet Found," 181.

[84] "I place myself squarely among those who believe that Könsgen went as far as due caution would allow in printing the *Epistolae* with the subtitle (and the punctuation thereof) he used, and that Mews and his supporters have gone too far." See ibid., 201.

[85] Ibid., 175.

[86] See Newman's review of the book in Barbara Newman, *00.01.06, Mews, The Lost Love Letters of Heloise and Abelard*. The Medieval Review, 2000, accessed October 6, 2019, `https://scholarworks.iu.edu/journals/index.php/tmr/article/view/14861`; for her criticism against ideology, I am referring here to Newman, "Authority," previously discussed on p. 245.

[87] von Moos, *Mittelalterforschung*.

of Saint-Leo at Toul (c. 1080–1150),[88] or those of Roscelin of Compiègne, who called her a "very wise girl" (*puella prudentissima*),[89] and, again, Peter the Venerable.[90]

Despite this, assembling a corpus which contains a 'clean' authorial signal from Heloise is hardly possible. Aside from her letters to Abelard (Epp. 2, 4 and 6), which we will continue to approach in this chapter as disputed, only one short letter addressed to Peter the Venerable may classify as an independent piece. It does not function within or combined with work affiliated to Abelard, and was written c. 1144 (two years after Abelard's death in 1142).[91] Aside from this letter, we also dispose of a preface letter to the forty-two questions on the Bible known as the *Problemata Heloissae*, in which each of her questions are submitted and paired up with Abelard's responses (*solutiones*).[92] Scholars have also suggested that the *dehortatio* against marriage in the *Historia Calamitatum* might be drawn up in Heloise's Latin,[93] but here again, influences from Abelard cannot be ruled out. This same concern holds true for the indirect citations and paraphrases ascribed to Heloise in Abelard's preface letter to the *Hymnarius Paraclitensis*. In the letter, Abelard cites Heloise, who solicits a new hymnal as the one in use at the Paraclete is severely corrupted by faulty transcriptions and dissenting versions, impossible to fit to a melody, and undermining the *auctoritas* of its supposed composers Jerome, Hilary and Ambrose.[94] Elsewhere, we have texts 'affiliated' with Heloise, which can be found in documentation on the early history of the Paraclete, as enlisted by Barrow, Burnett and Luscombe in their "checklist

[88] Letters 16 and 17 in Hugh Metel's collection. For further reference and an elaborate discussion of Hugh's appreciation of Heloise's literary style, see Constant J. Mews, "Hugh Metel, Heloise, and Peter Abelard: The Letters of an Augustinian Canon and the Challenge of Innovation in Twelfth-Century Lorraine," *Viator* 32 (2001): especially 76ff.

[89] As pointed out by Newman in Newman, "Authority," 128; for Roscelin's letter, see Roscelinus Compendiensis, "Der Brief Roscelins an Abälard", l. 10, 78.

[90] Especially in his first letter to Heloise (Ep. 115), to be discussed in n. 91 and written in response of a lost letter by Heloise. Here Peter testifies that "the fame, not yet of your piety, but of your distinguished and praiseworthy studies became known to me"("non quidem adhuc religionis tuae, sed honestorum tamen et laudabilium studiorum tuorum, michi fama innotuit"). He also refers to Heloise as "deeply devoted to literary studies, which is most unusual, and to the pursuit of wordly wisdom"("expeditam litteratoriae scientiae quod perrarum est, et studio licet saecularis sapientiae"). Translation taken from Petrus Abaelardus and Heloisa Argentoliensis, *The Letters of Heloise and Abelard: a Translation of their Collected Correspondence and Related Writings*, ed. and trans. Mary Martin McLaughlin and Bonnie Wheeler, The New Middle Ages (New York: Palgrave, 2009), 294.

[91] Peter the Venerable was Abelard's friend and protector, and he would procure Abelard's body for the Paraclete. Heloise's letter is contained within the letter collection of Peter the Venerable as Ep. 167, "Epistola Heloisae abbatissae ad dominum abbatem," see Constable's *The Letters of Peter the Venerable*, 1:400–1. Ep. 167 forms a response to a letter of Peter addressed to Heloise, namely Ep. 115 ("Ad Eloysam abbatissam," 303–8), in which Peter informs her on Abelard's death, and Ep. 168 ("Rescriptum domini abbatis," 401–2). In Ep. 167, Heloise solicits Peter's aid in providing a prebend for her and Abelard's son, Astralabe.

[92] Petrus Abaelardus and Heloisa Argentoliensis, *Problemata Heloisae*, in *PL* 178:677–730, ed. André Duchesne.

[93] The *dehortatio* —or *dissuasio*— heavily borrows from Jerome's treatise *Adversus Jovinianum*. See Ep. 1 Petrus Abaelardus and Heloisa Argentoliensis, *Letter Collection*, §24–7, 35–43; Silvestre and Fraioli have commented on the Heloise's misusage of the moral tenure of Jerome's treatise, wherefore they found the passage indicative of the letter corpus's inauthenticity. See Fraioli, "Importance of Satire"; and in that same volume Silvestre, "Die Liebesgeschichte."

[94] The *praefatio* is edited in Chrysogonus Waddell, ed., *Hymn Collections from the Paraclete*, Cistercian Liturgy Series, 8–9 (Gethsemani Abbey, Trappist, KY: Cistercian Publications, 1989), 5–9.

of manuscripts."[95] These include the liturgical ordinal of the Paraclete, a collection of Old French letters which were forged in the fourth- or fifteenth century and attributed to Heloise —*Les epistres de l'abesse Heloys du Paraclit*—,[96] a necrology of the monastery, and a number of contemporary charters by high-ranking clergy and nobility collected in the Paraclete cartulary. Amongst these documents, also the short customary for the Paraclete, the *Institutiones nostrae*, can be mentioned (discussed earlier on p. 241).

Mews's attribution of the text snippets by the *M<ulier>* in the *EDA* to Heloise invoked a series of additional ascriptions of anonymous poems and sequences to her, some more credible than others. In that same *Lost Love Letters* edition of 1999, Mews had already suggested two poems to have been the work of Heloise, because they were reminiscent of the *M<ulier>*'s style in the *EDA*.[97] He also suggested that a number of musical compositions (sequences), which could be traced back to the Paraclete and had originally been attributed to Abelard in 1986 by Waddell, could well have been Heloise's work instead, causing a sequel of scholarly debate with David Wulstan as the theory's defender and Dronke as its skeptical opponent.[98] Juanita Feros Ruys also came to attribute one of the *Carmina Burana* poems to Heloise (CB 126, *Huc usque, me miseram*) in a contribution called "Hearing Mediaeval Voices."[99] Despite

[95] Barrow, Burnett, and Luscombe, "Checklist," see especially 283ff.

[96] Also discussed in Dronke, *Medieval Testimonies*, 29ff.

[97] The first is a poem for the preacher Vital, founder of Savigny († 1122). The poem appears in a mortuary roll dedicated to Vital, and was written by a nun of Argenteuil in 1122, wherefore Mews jumped to the conclusion that it was to be attributed to Heloise who stayed there at the time. "Given that no other poet is known to have being living at Argenteuil in 1122, there seems little reason to doubt that Heloise is its author." See Mews, *The Lost Love Letters*, see 161ff, quote taken from 163; the poem was edited in Léopold Delisle, ed., *Rouleaux des morts du IXe au XVe siècle*, Société de l'histoire de France. Publications in-8 135 (Paris: Mme. Ve. J. Renouard, 1866), the mortuary roll as a whole is edited on 281–344, and the poem (counting some fourteen verses) is found on 299. The second is an anonymous poem from a Bury of Saint Edmunds verse anthology, edited in André Boutemy, "Le recueil poétique du manuscrit Additional 24199 du British Museum," *Latomus* 2, no. 1 (1938): see 42–4. The poem is written from the perspective of a woman writer lamenting that she is being forbidden to write. Concluding that the original author was female, and meaning to spot correspondences in the *M<ulier>*'s style once again, Mews raises the suggestion of Heloise's authorship.

[98] The three sequences are titled *Epithalamica*, *De profundis ad te clamantium* and *Virgines castae*. See Chrysogonus Waddell, "*Epithalamica*: An Easter Sequence by Peter Abelard," *The Musical Quarterly* 72, no. 2 (1986): 239–71. In 2002 and 2003, Wulstan followed Mews's lead and gave further arguments for their attribution to Heloise. In addition, he also attributed one of the *Carmina Burana* poems to her and two Catalan plays written by a dramatist from Vic. The *Carmina Burana* poem is called *Iam dudum estivalia* (CB 3 *), and the Catalan plays are titled *Verses de tres Maries* and *Versus de pelegrino*. For Wulstan's arguments to attribute the texts to Heloise, see David Wulstan, "*Novi modulaminis melos*: the Music of Heloise and Abelard," *Plainsong and Medieval Music* 11, no. 1 (2002): 1–23. And David Wulstan, "Heloise at Argenteuil and the Paraclete," in *The Poetic and Musical Legacy of Heloise and Abelard*, ed. Marc Stewart and David Wulstan (Ottawa: Institute of Mediaeval Music, 2003), 67–90. However, Dronke, who edited the plays and had indicated the similarities between the texts of which Wulstan had argued to be of same authorship, was quick to point out that Wulstan's attributions were based on rather careless assumptions and embarrassing misreadings of the Latin, and that the latter was "not adducing evidence but writing his own novelette."Dronke remained skeptical both of Waddell's ascriptions as those of Wulstan concerning the sequences' relationship to either Heloise or Abelard. Dronke's criticisms —and the 'novelette'-quote— can be found in Dronke and Orlandi, "New Works by Abelard and Heloise?," 131. For the edited text of the Catalan plays, see Peter Dronke, ed. and trans., *Nine Medieval Latin Plays*, Cambridge Medieval Classics 1 (Cambridge, UK: Cambridge University Press, 1994), 83–109. Dronke noted in ibid., 85 that the Easter sequence has parallels with the *Epithalamica*.

[99] Juanita Feros Ruys, "Hearing Mediaeval Voices: Heloise and *Carmina Burana* 126," in *The Poetic and Musical*

the alluring hypotheses formulated by all of these scholars, not hard evidence but a great deal of wishful thinking underlied the ascription to Heloise.

## 8.4 Consensus ex machina

### 8.4.1 Statistical Attribution Thusfar

An interesting trend in the Heloise-Abelard debate is that historians and literary scholars have more intensively sought for ways to discharge their personal intuitions, or distance themselves from using the kind of persuasive rhetoric that has occasionally done more damage than good. So, for instance, Jaeger, in his latest attempt in 2014 to argue for the *EDA*'s ascription to Heloise and Abelard, chose to "juxtapose passages from the *EDA* and the Abelardian/Heloisian corpus with minimal narrative connections and without argumentation. All debate is relegated to the notes, and all reference to arguments against the ascription occurs there."[100] In her 2016 translation and commentary of the *EDA*, Newman followed a similar type of 'demonstration without explanation.'[101] Evidence for this pursuit of neutrality can also be seen in scholars' frequent attempts to tackle the authorship question by applying statistical methods. As opposed to most of the other case studies presented in this thesis, the attribution dispute over Heloise and Abelard's letter corpus and the *EDA* already boasts quite an impressive record of stylometric experiments.

As early as 1914, Bernhard Schmeidler attempted to prove the stylistic unity of the letter corpus by analyzing such words and phrases as *tanto ... quanto*, *saltem*, and *obsecro*, reaching the conclusion that both Heloise's and Abelard's letters were written by Abelard alone.[102] In 1933, Charlotte Charrier came to similar conclusions.[103] Joseph Muckle's edition of the "Personal Letters" in 1953 also contained a section focused upon the *tanto ... quanto* construction, and likewise an analysis on the frequent use of the *ut*-relative clause at the end of a sentence, which the *Historia Calamitatum* had in common with the ensuing letters.[104] In 1975, John Benton —then still convinced that the letter corpus was a forgery—, responded to a request by M. Jacques Monfrin at the 1972 Cluny conference to perform a computer-assisted analysis,[105] and published a report with Fiorella Prosperetti Ercoli.[106] Inspired by Frederick Mosteller and David Wallace's analysis of function words (or 'style markers') in their ground-

*Legacy of Heloise and Abelard* (Ottawa: Institute of Mediaeval Music, 2003), 91–9.

[100] Jaeger, "An Annotated Concordance," 186.

[101] Newman, *Making Love*.

[102] Schmeidler, "Der Briefwechsel zwischen Abälard und Heloise eine Fälschung?"

[103] Charrier, *Héloise*, 582–90.

[104] Muckle, "The Personal Letters Between Abelard and Heloise," see especially 52–4.

[105] Benton, "A Reconsideration," 44.

[106] John F. Benton and Fiorella Prosperetti Ercoli, "The Style of the "Historia Calamitatum": A Preliminary Test of the Authenticity of the Correspondence Attributed to Abelard and Heloise," *Viator* 6 (1975): 59–86.

breaking study of 1964,[107] Benton came to the conclusion that the *Historia Calamitatum* was not of Abelard's writing, although he later confessed that "six words not even chosen at random are insufficient evidence on which to determine authorship," a justified concern.[108] In 1988, Tore Janson concluded from both Dronke's[109] as his own statistical experimentation with the cursus that the Heloise-Abelard collection was written by one and the same author, although he left open whether Heloise or Abelard was responsible.[110]

Von Moos provided a table of percentages indicating the favoured use of *cursus* prose cadences by both <*V*>*ir* as <*M*>*ulier* in the *EDA*.[111] Dronke and Orlandi applied $\chi^2$ (Pearson's chi-square) to the *EDA* and the letter collection (manually annotated) and likewise compared rhytmic cadences.[112] They came to the conclusion that the *EDA* are written by two individuals different than Heloise and Abelard: the <*M*>*ulier*'s application of it is much more frequent than that of <*V*>*ir*'s, but also than that of Heloise's. Ziolkowski, sympathetic to stylometric practices in impasses such as that of Heloise and Abelard, provided an additional table of manageable function word frequencies —*autem*, *igitur*, *ergo*, *ita*(*que*), *quia* and *quippe*— which further endorsed Dronke and Orlandi's findings, and showed that Abelard's and <*V*>*ir*'s writings were too different to be of same authorship.[113] Stella, in measuring lexical overlap of words, bigrams and trigrams, agreed with Ziolkowski, Dronke and Orlandi that the *EDA* are written by two distinct authors.[114]

## 8.4.2 New Stylometric Evidence: Do We Need It?

### Technological Advancements

Having reviewed this long list of former achievements, and having worked through the debate's long-standing history, an inevitable question arises: who is still waiting for the question of Heloise and Abelard's authorship to be answered, or —perish the thought— who is willing to discuss all over again the validity and invalidity of arguments for and against? Not surprisingly, some dust appears to have settled on

[107] The study is treated in more detail on p. 66 in this thesis. See Mosteller and Wallace, *Applied Bayesian*.

[108] Benton, "A Reconsideration," 44.

[109] Peter Dronke, "Heloise's *Problemata* and *Letters*: Some Questions of Form and Content," in *Petrus Abaelardus (1079–1142). Person, Werk und Wirkung*, ed. Rudolf Thomas, Trierer Theologische Studien, 38 (Trier: Paulinus Verlag, 1980), for the data, see the article's addenda on 71–3.

[110] Tore Janson, "Schools of Cursus in the Twelfth Century and the Letters of Heloise and Abelard," in *Retorica e poetica tra i secoli XII e XIV. Atti del secondo Convegno internazionale di studi dell'Associazione per il Medioevo e l'Umanesimo latini (AMUL) in onore e memoria di Ezio Franceschini, Trento e Rovereto 3–5 ottobre 1985*, ed. Claudio Leonardi and Enrico Menestò, Quaderni del Centro per il Collegamento degli Studi Medioevali e Umanistici nell'Universita di Perugia 18 (Perugia, Firenze: «La Nuova Italia» Editrice, 1988), 171–200.

[111] von Moos, "Die *Epistolae duorum amantium*," see "Anhang I: Cursus" on 103.

[112] The full scansion is available in Dronke and Orlandi, "New Works by Abelard and Heloise?"

[113] Ziolkowski, "Lost and Not Yet Found," 191.

[114] Stella, "Analisi informatiche."

the debate in the past years. Especially for the Heloise-Abelard collection, consensus has risen (perhaps from fatigue?) that the correspondence is genuine, written by the prioress and the founder of the Paraclete: Heloise and Abelard. For the *EDA*, such a consensus does not seem to have been established, although my impression is that the most commonly held opinion is situated in the spectrum 'agnostic—opposed.'

It has been forty-seven years since Benton's groundbreaking paper of 1972, and computational methods which were then still largely unknown, are now increasingly gaining leeway in the humanities. The days of manual counting are behind us. The computational power of current-day machines has drastically improved, and far-reaching textual similarities can be detected in a few milliseconds. Most stylometric research on Heloise and Abelard's works were carried out in the heat of the debate, at which point such technology was not yet available. Most of these statistical experiments are outdated, restricted in scope and non-exhaustive, performed by non-experts in the field of computational stylistics and unverified concerning methodological validity (can prose rhythm in fact be shown to systematically achieve high accuracy for authorship attribution in medieval Latin?).

The new stylometric evidence which I will bring forth will take into account these limitations and try to formulate an answer to them. As the former chapters will have made amply clear, distance and difference are relative concepts. This has not always been fully understood, or could not always be fully tested in previous stylometric experiments on the two letter collections. So even though I will start from an up-close, internal analysis of stylistic patterns, I will also come to verify these micro-patterns' relevance on a macro-level by bringing in external corpora by other twelfth-century authors, thereby gradually maintaining a view on the larger picture. As with other experiments in this thesis, I will also consistently attempt to illustrate how the method works and why it is valid to believe that what I am measuring is not simply 'some' stylistic difference, but a difference that can be traced back to dual, collaborative or single authorship. I believe, once more, that some confusion may have taken place amongst these categories in former research on the collections, especially for the *EDA*.

### The Collaboration of Heloise and Abelard

Aside from such technical considerations, I will come to relate the results below with the larger research questions of this dissertation, in particular the distribution of authorial roles, with a focus on cross-sex collaborative composition of Latin literature. After all, from preface letters to some of Abelard's extant works, and from text-internal clues in the Heloise-Abelard correspondence, we know that Abelard wrote many works at the request of Heloise and the nuns at the Paraclete, such as sermons, hymns (*Hymnarius Paraclitensis*), poems and commentaries (*Expositio in Hexameron*). For the *Problemata Heloissae* and the Heloise-Abelard letter exchange, whose compositions are conventionally dated in the 1130s, a close collaboration between

| Author | Title (or *incipit*) | Length |
|---|---|---|
| Peter Abelard | *Apologia contra Bernardum* | 2761 *w* |
| | *Commentaria in epistulam Pauli ad Romanos* | 79,990 *w* |
| | Epp. 9–14 | 12,484 *w* |
| | *Expositio in Hexameron* | 24,034 *w* |
| | *Glossae super Hermeneias* | 99,469 *w* |
| | *Scito te ipsum* | 18,206 *w* |
| | *Sermones* | 90,783 *w* |
| | *Sic et non* | 124,469 *w* |
| | *Theologia Christiana* | 85,658 *w* |
| | *Theologia Scholarium* | 62,425 *w* |
| | *Theologia 'Summi boni'* | 29,088 *w* |
| Heloise | *Epistola ad Petrum Venerabilem* | 219 *w* |
| | *Epistola Problemata Heloissae* | 437 *w* |
| | *Praefatio hymnorum* | 592 *w* |
| | *Institutiones nostrae* (?) | 1,121 *w* |
| *EDA* | <M>*ulier* | 6,544 *w* |
| | <*V*>*ir* | 5,930 *w* |

Table 8.1: Corpus for Heloise and Abelard, and word totals for the *EDA*'s <M>*ulier* and <*V*>*ir*. The *Institutiones nostrae* are listed with Heloise's works, although I do not necessarily subscribe to Waddell's ascription (1987) until further evidence is provided.

Heloise and Abelard seems feasible, considering how the works are constructed from a question-response formatted interaction between them. Obviously, if the collection is to be held authentic, Heloise played an important role in its composition. My question is therefore not so much aimed at either aggrandizing or downplaying that role, but to assess by computational means the degree by which we may arrive at a consensus upon authenticity or inauthenticity, upon attributing to Heloise what may be attributed to Heloise, and of gaining a better intuition of the dynamics at work when Heloise and Abelard composed their letter collection.

## 8.5 Experimental Set-Up

### 8.5.1 Corpus

Table 8.1 enlists the corpus we will be working with in the coming experiments. Taking into account the word totals in the right column, the minimal sample length for the experiments below was set at 1,100 *w* (the length of the *Institutiones nostrae*). This minimum sample length is compatible with the minimum 'reliable' number of written words we have for Heloise, which is 1,200 *w*. This number is reached when adding up the words of the preface letter to the *Problemata Heloissae*, the indirect citations in the preface letter to Abelard's hymnal (Abelard's sentences in the preface could obviously not be included) and the words from her letter to Peter the Venerable.

As noted above on p. 249, only the letter to Peter the Venerable can stand as a piece certainly not influenced by Abelard.

All in all, the outlook of a corpus such as this one is not dramatic per se. By now we have seen worse than 1,100 *w* as a minimum sample length. Nevertheless, the lack of reliable training data makes a supervised approach —where we can distinguish between classes and train on them— very difficult, if not impossible. If we stand by the minimum sample length of 1,100 *w* determined by the *Institutiones nostrae* (which only seems advisable), we would only be able to train an algorithm on a single sample of text for Heloise. This is impossible, since in order to demarcate classes, we need development data to evaluate the algorithm's progress.

## 8.6 Results

### 8.6.1 *Letter Collection*

What we see in fig. 8.2 is a pair of three-dimensional scatter plots, each of which visualizing Heloise and Abelard's letter collection (the *EDA* are not yet included). Also included in this scatter plot are the *Institutiones nostrae*, tentatively attributed by Waddell to Heloise in 1987.[115] The three short letters of Heloise[116] have been combined into a single sample and form a total of 1,100 *w* (green star). Abelard's Epp. 9–13, edited by Edmé R. Smits, have also been included (red stars).[117] The colours indicate the state-of-the-art ascriptions, Abelard in red and Heloise in green. The digits overlapping the plot markers (to be consulted in the legend) indicate where possible the letters' index in the editions of Luscombe and Smits (not the sample index). Subplots 8.2a and 8.2b differ from one another by their feature subsets, namely most frequent words (MFW, *n*=1,000) vs. function words only (*n*=350).[118] Note that MFW includes lexical items that may capture overlap in content, whereas function words capture syntactic and grammatical patterns of style (the full list of function words is given in the addenda to this chapter from p. 368 onward).

#### Date and Context

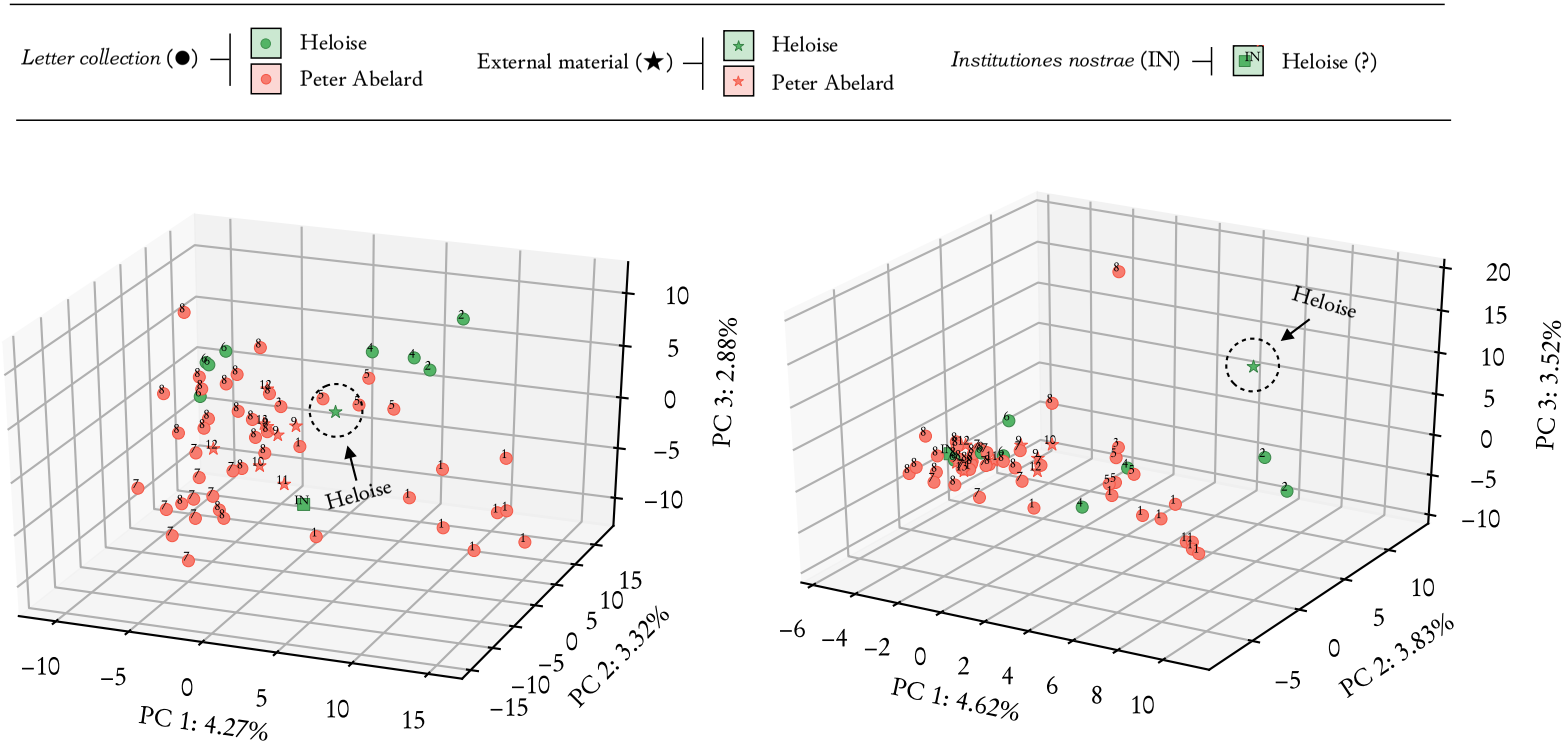
Firstly, having Abelard's external letters in the mix (Epp. 9–13, star-marked red samples) is steadily making the idea that the letter collection is not a twelfth-century composition from within the Paraclete milieu untenable. Epp. 9–13 form a set of external

[115] Waddell, *The Paraclete Statutes*.

[116] The preface letter to the *Problemata*, the preface letter to the *Hymnarius Paraclitensis* and the letter to Peter the Venerable.

[117] Ep. 14 is too short (439 *w*) and was excluded. See Petrus Abaelardus, *Peter Abelard. Letters IX-XIV*, ed. Edmé Renno Smits (Groningen: Rijksuniversiteit, 1983).

[118] The number of function words is limited. At a certain point, one will exhaust the maximum number of extant function words for Latin, hence the difference in number of features with MFW (*n*=1,000). The total list of function words used in generating this figure is given in the appendix on p. 368.




(b) MFW

(b) Function words

Figure 8.2: PCA plots of the Heloise-Abelard collection, combined with external letters by Abelard (Epp. 9–13) and Heloise's external letters. The digits overlapping the plot markers indicate where possible the letters' index (e.g. 1 = Ep. 1). Settings of plot fig. 8.2a: *s–l* = 1,100 *w* | *type* = most-frequent words | *n* = 1,000 | *vect.* = standard-scaled tfidf-weighted frequencies. Settings of plot fig. 8.2b: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies. Complete list of function words used in generating fig. 8.2b is given on p. 368.

letters by Peter Abelard, which were not transmitted in the collection's dominant Troyes manuscript, but survive in various and predominantly late manuscripts (i.e. fifteenth–seventeenth centuries) as well as in the 1616 principal edition by d'Amboise and Duchesne (*PL* 178:325–55).[119] One of the letters, however, Ep. 11, survives in two twelfth-century manuscripts.[120] Considering these external letters' strong affiliations with the Heloise-Abelard letter collection, the collection must —a point upon which the majority of scholars had agreed by now— be considered a twelfth-century text, and one in which Abelard himself was definitely involved. That is, unless we presume that the forger of the collection possessed of the skill to imitate Peter Abelard by the function word (350 function words, to be exact), or if we were to maintain that all of Abelard's letters are the work of the same forger. None of both seem very likely.

## Authorship

The collection's authorship by Heloise and/or Abelard, however, is another matter. As can be inspected in fig. 8.2, very few distinct clusters are produced in the PCA, and our pre-annotated colours hardly match with any observable trend in the data. Especially Heloise's Epp. 4 and 6 and Abelard's Epp. 1, 3, 5, 7 and 8 are not easily distinguished. Although a straightforward answer cannot be given, these first two plots suggest that single authorship of the collection is most likely. Here is why:

1. As already touched upon, the letter collection as a whole shows strong similarities with Abelard's external letters (Epp. 9–13, red stars).

2. Heloise's external 'corpus,' which forms only a single, green star-marked sample, drifts off from most of the other samples when scanned for function word distributions. When analyzed by MFW, all works cluster together, and no distinctions are produced (this includes the *Institutiones nostrae*, which cannot unambiguously be related to any of the samples traditionally presumed to be the work of Heloise).

3. One important exception requires further study: subplot 8.2b, in which the set of letters is explored by using function word distributions, suggests that Heloise's Ep. 2 has similarities to the Heloise as we know her from the external letters. We will soon come to discuss, however, that this visual impression at first glance is best approached with due caution.

Especially the last point, the behaviour of Ep. 2, deserves closer scrutiny. In fig. 8.3, we revisit fig. 8.2b, and its PC's are recast to a pair of two-dimensional PCA plots,

[119] The manuscripts used for this edition are discussed in n. 29.

[120] Paris, Bibliothèque nationale, lat. 356 and ibid., lat. 2445A. See Barrow, Burnett, and Luscombe, "Checklist," 212–3.

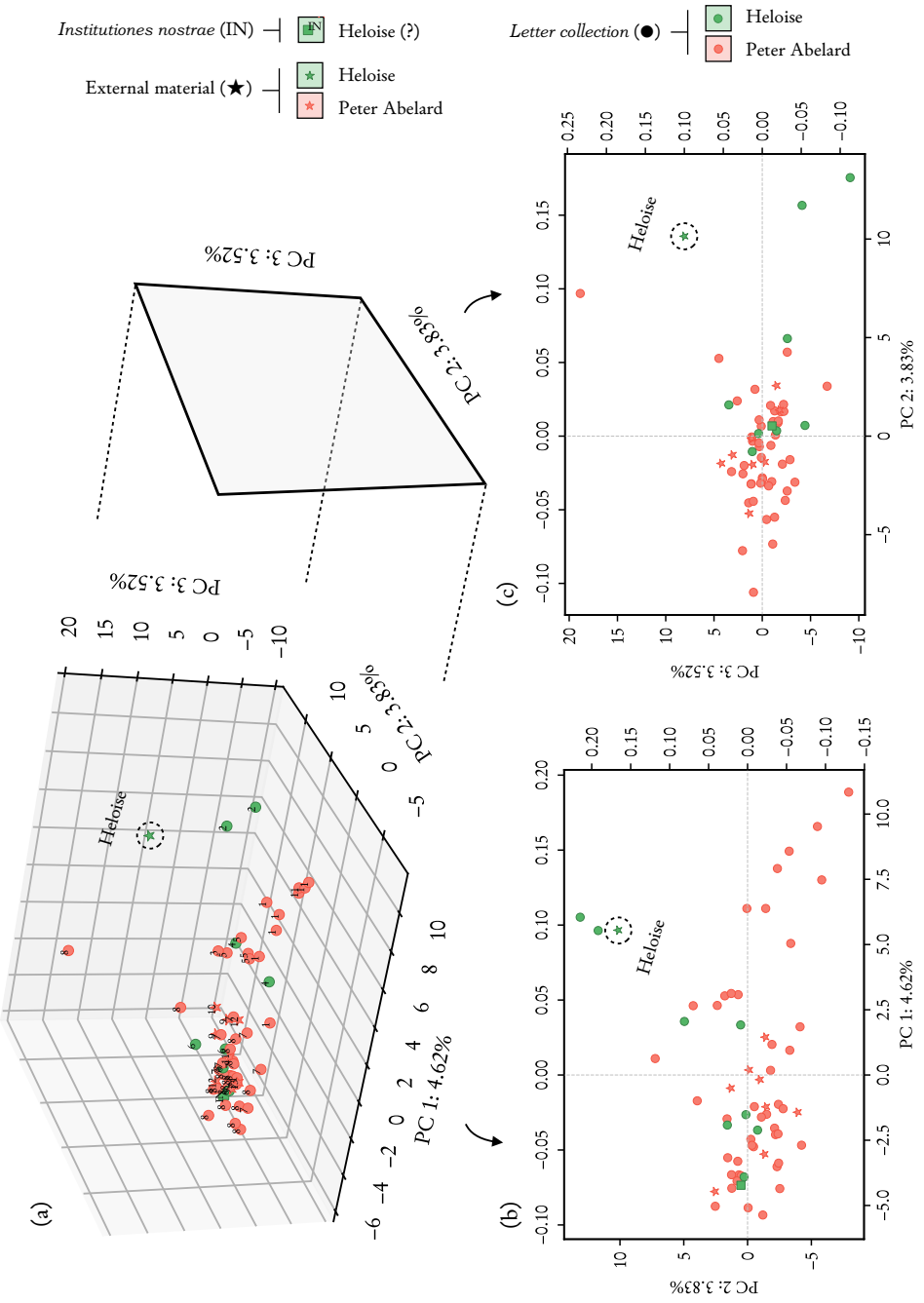



Figure 8.3: Rehearsal of fig. 8.2b, where the three-dimensional visualization is recast into two two-dimensional PCA plots. Settings of all three plots: $s$–$l$ = 1,100 $w$ | *type* = most-frequent function words | $n$ = 350 | *vect.* = standard-scaled tfidf-weighted frequencies.

which when viewed in complementarity offer a very nuanced intuition of Ep. 2 being the work of Heloise. Although fig. 8.3b suggests that Heloise's external works (green star) have quite a striking affiliation to this letter, fig. 8.3c shows that they are also drastically different when the third PC is introduced. We will soon present additional evidence for arguing that Ep. 2 displays 'outlying' behaviour which is dissimilar from all of the other writing samples in the collection.

To place the findings of figs. 8.2–8.3 in perspective, fig. 8.4 introduces two new subplots, in which Heloise and Abelard's letter collection is no longer analyzed in isolation, but is benchmarked against additional texts. In subplot 8.4a, Hildegard of Bingen and Bernard of Clairvaux's respective *epistolaria* are introduced (only outgoing correspondence). This immediately nuances some of the 'distances' as they appeared earlier. Heloise's external works (green, star-marked sample) generally keeps very close to the stylistic profile of the Heloise-Abelard collection. A problem here is that PCA is only advisable for a set of three authors: at this point it is difficult to verify if this condition is still respected, since the total number of authors involved may be higher than three.[121]

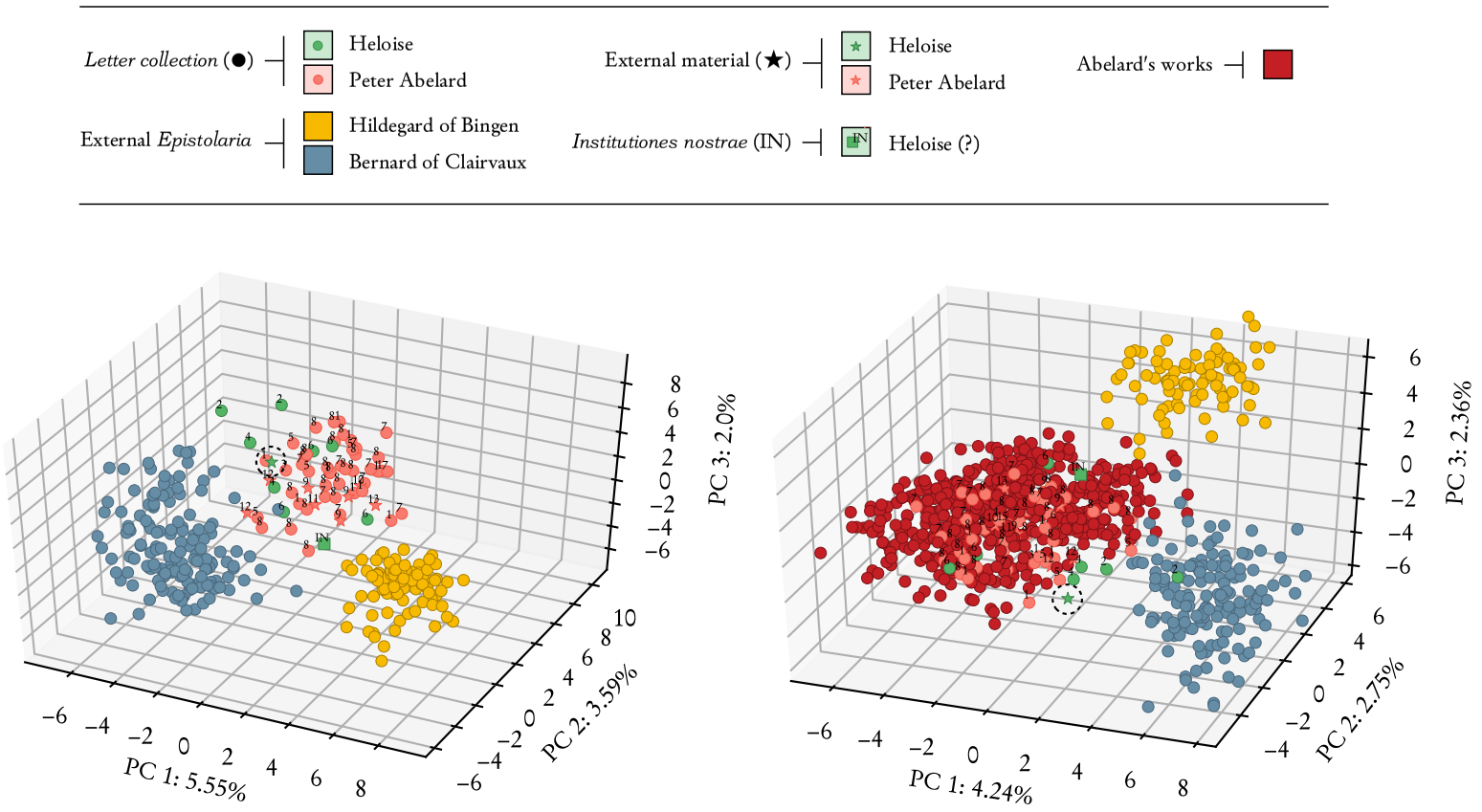
In subplot 8.4b, we benchmarked all of Abelard's works aside from his letters (listed earlier in table 8.1) against the texts that were contained in fig. 8.4a. One can observe that Abelard's works mingle freely with samples from the letter collection, including those traditionally ascribed to Heloise. One may perhaps argue that Epp. 2 and 4 and Heloise's external works set themselves somewhat apart, but the general trend remains that the bulk of text which scholarship currently ascribes to Heloise and Abelard seems to consist of very homogeneous and near to inextricable material.

Again, in both of these plots, Ep. 2 behaves most curiously, and appears to seek some detachment from the Heloise-Abelard collection. Ep. 2's distancing from the collection could be meaningful, but it is hard to defend this divergence as evidence of authorship by Heloise alone, given that there is no convincing resemblance to Heloise's external letters (green star). Combining the evidence from fig. 8.3b and figs. 8.4a–b, I can only conclude that Ep. 2 displays quite strong 'anomalous behaviour' as opposed to the other samples in the letter collection. It remains only the question what the underlying cause of this anomalous behaviour is. We will return to this point below.

### *Epistolaria*

One may rightfully bring up the objection, already formulated some years ago by Piron in response to calculations made by Orlandi, Stella, Benton and Ziolkowski for the Heloise-Abelard collection and the *EDA*, that strong agreement of word frequencies does not measure same authorship, but rather an agreement between two correspondents who "tend to use similar vocabularies":

[121] Binongo and Smith, "The Application of Principal Components Analysis to Stylometry," 464.




(a) Letter collection benchmarked against the *Epistolaria* of Hildegard of Bingen and Bernard of Clairvaux.

(b) Letter collection benchmarked against Peter Abelard's canonical works.

Figure 8.4: PCA plots of the Heloise-Abelard collection. Fig. 8.3a visualizes the collection in combination with additional *Epistolaria* of Hildegard of Bingen and Bernard of Clairvaux. Fig. 8.3b visualizes the collection in combination with Abelard's works as listed in table 8.1. Settings of both plots: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies.

> Out of my own calculations on the vocabulary of both collections, I observe some notable proximities between the woman and the man in the *EDA*, as well as between Heloise and Abelard in the Correspondence (precisely those signs that prompted John Benton to argue for a single author theory in 1975). Yet this shows primarily that people writing to each other tend to use similar vocabularies. The frequency of many function words in the woman's and Heloise's letters is rather close (for instance, *enim*, *ergo*, *iam*, *ita*, *nunc*, *pro*, *sic*, *tam*, *unde*), and the mutual overlap of their vocabulary is notable, but certainly not strong enough to claim an identity of authorship on that sole basis.[122]

Piron makes a good point. In fact, one can take this criticism further. The Heloise-Abelard collection is a carefully composed, single *epistolarium* with a clear literary purpose. These are not merely individual *epistolae* sent to and from, influencing two correspondents' linguistic characteristics. These letters have been assembled (and were maybe intended from their very inception) to be coherent and form a narrative unit, which is true for almost all twelfth-century letter collections. What exactly must be understood from "similar" styles and "close" frequencies, and what distinguishes "notable" from "not strong enough" —in Piron's quote— is bound to remain relative, however, unless we try to test Piron's assumptions and learn from inductive experience.

As has been noted on many occasions by Constable,[123] one is better at guard when assessing the authorship(s) of incoming and outgoing correspondence if the pieces have been united in single corpus, which is a hypothetical course of affairs for the composition of the Heloise-Abelard collection as well. It was customary for letter collections to receive a final polishing, or in some cases one may go as far as to say that their contents were completely fabricated for the sake of an enticing life story: also the incoming letters by various authors could receive their share of revision. Given the idea that incoming and outgoing correspondence was often contrived, and that stylistic anomalies were smoothened out in the process, the question might justly rise if the stylistic homogeneity, as it is expressed for the Heloise-Abelard collection above, forms an exception at all in medieval *epistolaria*. In chapter 4, we have already addressed this issue by showing that Nicholas of Montiéramey's stylistic influence is predominantly visible in those letters which never made it to Bernard of Clairvaux's official and public collection.[124] By contrast, Bernard's stylistic mark rests heavily on those letters which he intended to be read by a wider audience.

Admittedly, finding perfectly comparable situations to a unique corpus such as the Heloise-Abelard collection is not possible, and can only be selected by approximation.[125] What makes the exchange unique is that it presents itself as an exclusive exchange between two correspondents, whereas for most other (larger) *epistolaria*, in-

[122] Piron, "Heloise's Literary Self-Fashioning," 132.

[123] Constable, *The Letters of Peter the Venerable*, especially 2:1–44; and Constable, *Letters and Letter-Collections*.

[124] I am referring to the difference between the *extra* in the *intra* corpus, explained in chap. 4 on p. 131 above.

[125] "The survival of a true epistolary exchange between two medieval personalities, i.e. of a closed collection of letters

coming correspondence was seldom included at all, and if it was, the author took care that multiple correspondents were involved so as to enhance the authority of his or her persona. In what follows, we propose two lengthy *epistolaria* where mutual influence is possible, or in which we expect that one of both correspondents may have revised the style of the collection for proper benefits. This enables us to place the interpretations of the Heloise-Abelard collection (especially the single authorship hypothesis) in a somewhat wider scope.

A first and well-known example is the epistolary collection of Peter the Venerable containing 142,262 *w* of *epistolae* written by Peter himself, and twenty-four incoming letters (9,360 *w*).[126] Constable edited and printed the letters of Peter the Venerable as faithfully as possible when it comes to their appearance and arrangement in the original collection. The very few individual pieces surviving in copies outside of the official collection, in a recipient's archive or in variant witnesses, have indicated how the final collection was, however, "for some twenty-five years [...] a changing, living text."[127] Peter's *epistolarium* clearly received an intensive editorial revision prior to its publication, and efforts were not limited merely to those letters of Peter's authorship. As can be deduced from the variant versions given in Constable's *apparatus*, the incoming correspondence by other authors may have seen its share of stylisic emendations. The extent of the stylistic alterations (that we can objectively witness through comparing different copies) varies from extensive to negligible, but in either cases there can rarely be any certainty of what the original missive looked like before it entered the collection.

The second collection of interest is that of Hildegard of Bingen, which not solely contains Hildegard's outgoing letters (316 letters, 96,795 *w*) but also those incoming from various correspondents (138 letters, 25,971 *w*).[128] As touched upon in chapter 6, Hildegard's *epistolarium* was a carefully planned composition, for which the first outlines were drafted still during her lifetime with the assistance of Volmar. After the latter's death († 1173), the older letters were adopted, revised and together with new letters (first composed after 1173) appended to her Riesencodex by Guibert of Gembloux.[129] When it comes to the stylistic outlook of the incoming letters, the same

of which each letter is the answer to the preceding one and the stimulus for the succeeding one, is a unique feature in the field of text transmission. Letter collections of all times are characterised by their unilateral perspective: they contain only outgoing letters with exceptionally an incoming one. Even in the didactic tools, the *Artes dictaminis* with their model letters, a true exchange is rarely represented," see Verbaal, "Epistolary Voices," 11.

[126] Aside from the first letter of the collection (*praemissa*) by secretary Peter of Poitiers, the incoming letters in Peter the Venerable's collection are Epp. 25, 41, 62, 71, 78, 85, 96, 114, 123, 125–8, 146, 153–5, 158b, 165, 167, 169, 179, 187.

[127] See Constable, *The Letters of Peter the Venerable*, 2:44.

[128] The *epistolarium* is edited in three volumes in Hildegardis Bingensis. *Epistolarium*. Ed. by Lieven Van Acker and Monica Klaes-Hachmöller. *CC CM* 91–91B. Turnhout: Brepols, 1991–2001. The incoming letters are Epp. 4, 10, 13–6, 18, 20, 22, 24–7, 29, 31–2, 34–5, 37–41, 43–5, 48–53, 55, 60–2, 66, 68–70, 72, 74, 76–8, 80, 82–5, 87, 91, 93–4, 97–8, 100, 110–5, 117, 120–1, 138–40, 142, 144, 146–50, 153–60, 163–70, 172–4, 176–79R, 182–3, 185–8, 190–1, 195, 200, 204, 206–7, 209, 215, 220, 221, 223–4, 227–31, 237, 240–1, 244, 250, 265, 269–70, 276, 295–7, 311, 314, 324.

[129] For the Riesencodex, see p. 179 earlier in this thesis. Van Acker established that the *epistolarium* in the Riesencodex

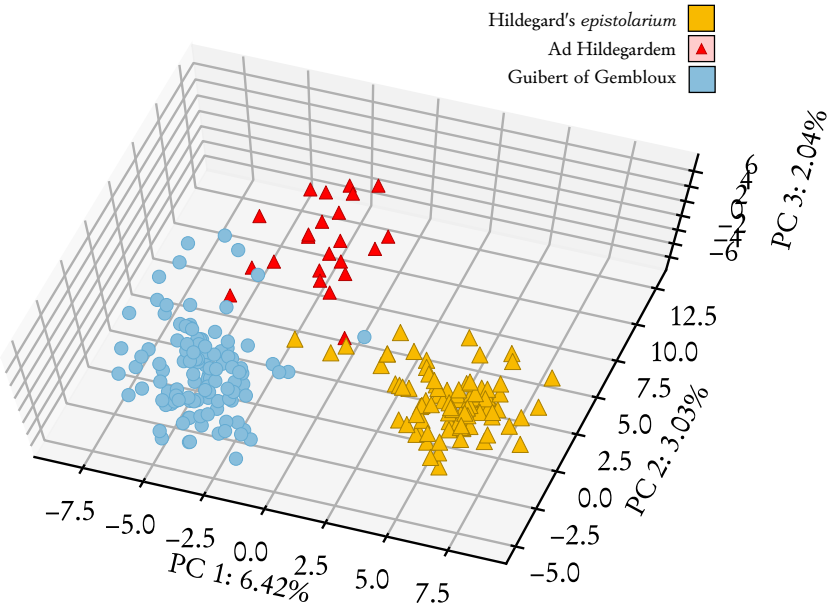



Figure 8.5: PCA plot of Hildegard of Bingen's *epistolarium*, both outgoing (yellow) as incoming (red) correspondence. Also the *epistolarium* of Hildegard's last secretary Guibert of Gembloux (in blue) is included. Settings: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies.

procedure of revision —or downright fabrication?—[130] as with Peter the Venerable may have taken place. Aside from a few exceptions, such as letters from Guibert of Gembloux, Elisabeth or Schönau and Bernard of Clairvaux (who have a copy of their work sent to Hildegard included in their own collections), we do not possess alternative redactions of the incoming letters in archives or copybooks to verify if they are authentic or heavily revised by Hildegard or her entourage, principally Volmar and Guibert.

The results of the stylistic analyses on these two letter corpora by Peter the Venerable and Hildegard of Bingen, displayed in figs. 8.5–8.6 are striking, and could not have been more different.

Let us first discuss the most straightforward of both collections, namely that of Hildegard of Bingen in fig. 8.5, which can immediately be placed in contrast with

(*R*) (MS Wiesbaden, Nassauische Landesbibliothek, Cod. 2) is a later transcription (and revision) of a lost antecedent *Wr(u)*, which contained Hildegard's letter collection before 1173 (the date of Volmar's death). What this twelfth-century non-existent *Wr(u)* looked like is recoverable only by a mid-thirteenth-century witness MS Vienna, Österreichische Nationalbibliothek, Cod. 963 (theol. 348), which goes by the siglum *Wr*. The manuscript is younger than *R*, but contains an older attestation of Hildegard's letter collection. See Lieven Van Acker, "Der Briefwechsel der heiligen Hildegard von Bingen. Vorbemerkungen zu einer Kritischen Edition," *Revue Bénédictine* 98, nos. 1–2 (1988): especially 152. Those letters written before 1173 —under Volmar's redaction— and those after 1173 have been shown to have a different stylistic profile. See Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 213–4.

[130] For the forging of the *epistolarium*'s original first letter, see p. 180 earlier in this thesis. For the *epistolarium*'s purpose to represent Hildegard's persona, see Van Engen, "Letters and the Public Persona of Hildegard."

the Heloise-Abelard collection. None of Hildegard's correspondents appear to have adjusted to her language (red triangles), nor the other way around.[131] Also Guibert of Gembloux did not appear to have meddled with the text of the incoming letters. Importantly, this short sidetracking into Hildegard's *epistolarium* shows that personal stylistic profiles can be conserved within a single manuscript, in carefully constructed *epistolaria* which have seen their share of revision.

Something very different happens in the letter collection of Peter the Venerable. At first glance, fig. 8.6, which visualizes the collection (with colours annotating which samples belong to the outgoing and incoming correspondence in respectively blue and red) renders the collection surprisingly homogeneous, where even incoming correspondence converges with the letter corpus's overall style. Must we conclude that Peter the Venerable adjusted the style of the incoming correspondence as drastically as it appears?[132] There is no easy answer, and we will have to discuss some additional movements that do not immediately meet the eye, and are stirring behind the surface.

If one glances more closely at fig. 8.6a, one will see a group of loose data points (blue triangles) swarming on the right side of Peter the Venerable's cluster of letters in the collection, with large interlying spaces. These are all samples belonging to Ep. 28, Peter's monumental, popular treatise in epistolary form addressed to Bernard of Clairvaux, which had the intention of defending Cluniac traditions against those of the Cistercians. The text is well-known to boast a complex and various transmission history, and was revised over and over, both during Peter's lifetime as after.[133] Constable has dedicated a number of pages in his introduction to the collection of the exceptional letter, commenting on the letter's various guises: "it is clear that there was no fixed and authoritative form of the text." It was a continual "work of revision" which "probably began as soon as the first copies had left the *scriptorium* at Cluny."[134]

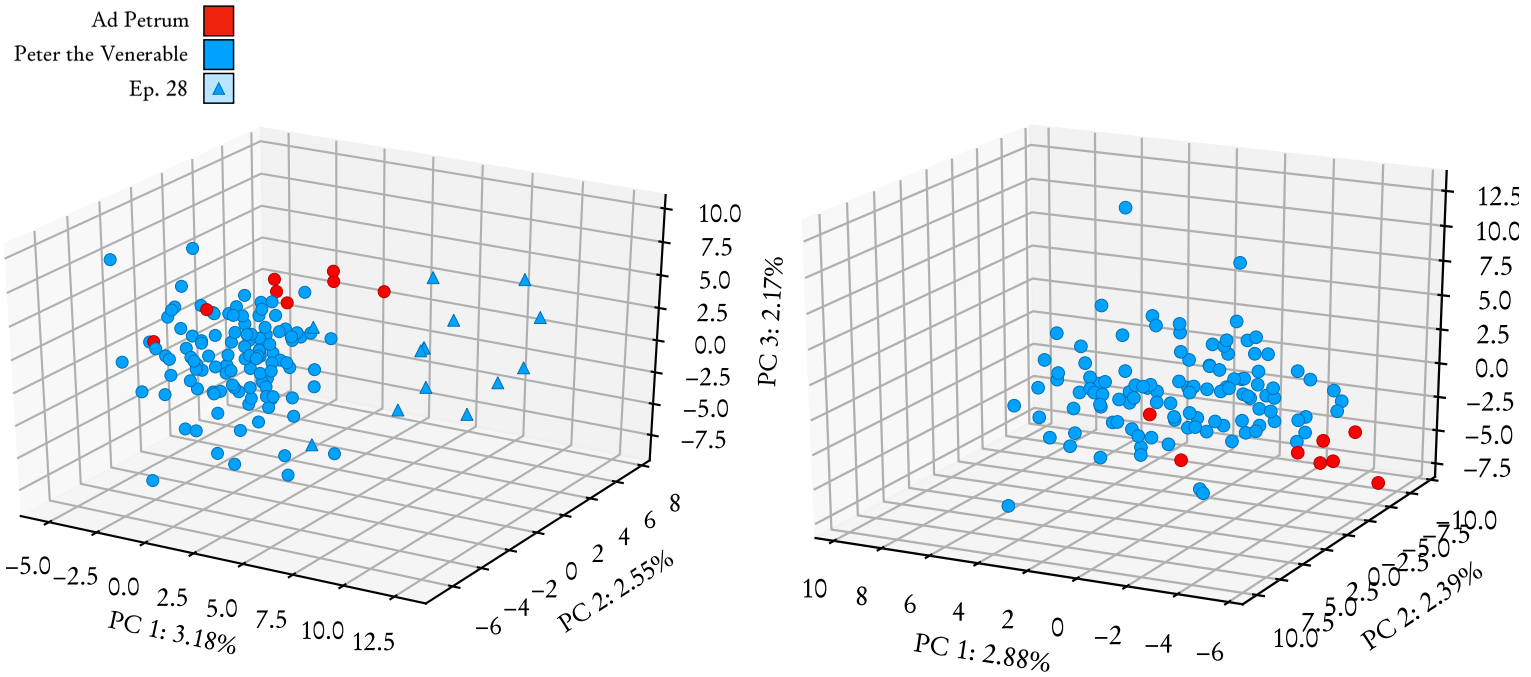
It is important to bring home the following point: the 'principal' stylistic difference within Peter's collection as picked up by PCA does not seem to be brought about by the collection's inclusion of incoming letters, which may have been severely revised, but by Ep. 28. One could go as far as to say that Ep. 28 is more different from Peter's own letters, than Peter's own letters are different from other authors. This raises an important issue, which is at once methodological, appertains to the case study of this chapter, and revisits the general question of collaborative authorship central in this thesis: which of these samples offers us Peter the Venerable, and which of them represent the copyists, revisors, secretaries, or any synergetic combination of such

[131] Note that Hildegard's various correspondents (in red) only appear to cluster together because the data's 'principal trends' —which is what PCA attempts to find— is here obviously the difference between what is Hildegardian vs. what is not-Hildegardian: there are too many authors visualized in the PCA in order to gain a full-dimensional account of what segregates the samples of all the individual writers other than Hildegard. By contrast, there appears to be no such principal difference between Heloise and Abelard.

[132] Waddell, *Hymn Collections*.

[133] See Constable, *The Letters of Peter the Venerable*, 2:63–70.

[134] Constable, *The Letters of Peter the Venerable*, 2:68.




(a) Collection with Ep. 28.

(b) Collection without Ep. 28.

Figure 8.6: PCA plots of Peter the Venerable's *epistolarium*, both outgoing (blue) as incoming (red) correspondence. The left-hand plot depicts the *epistolarium* including Ep. 28, Peter's best known letter in defence of Cluniac traditions, and the right-hand plot visualizes the collection without. Settings are identical for both plots: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies.

actors? It should also be noted that this find —which suggests a very heavy revision process— could be detrimental for the authenticity of the only letter we have for Heloise that existed independently of Abelard: Ep. 167 was, after all, one of the incoming letters in Peter the Venerable's *epistolarium*. The letter in itself is too short (213 *w*), however, to make a meaningful assessment of how Peter the Venerable may have meddled with Heloise's words.

The experiments with these additional *epistolaria* have indicated once more that 'distance' and 'variation' are relative givens, and that it may be instructive and even advisable to compare similar corpora with one another. Fig. 8.7 paints a clear picture of all *epistolaria* in one and the same plot, and shows once more just how deep-rooted the stylistic similarity of the letters ascribed to Heloise and Abelard is. Without sidetracking here on the questions of gendered authority and synergy that arise from the results of Peter the Venerable and Hildegard of Bingen's *epistolaria*, which deserve a study on their own, I will for now relate these findings to the Heloise-Abelard collection, and formulate a number of careful conclusions.

We have seen two additional examples of *epistolaria*. The first, that of Hildegard of Bingen, gave indication that individual letters in letter collections may still contain personalized forms of expression. The second, however, of Peter the Venerable, which testifies of a high degree of collaboration and revision, presents us with a case in which the stylistic profile of correspondents seems to have been effaced up to the point of becoming unrecognizable. However, when considering the homogeneity of Peter the Venerable's letter collection, one will notice that here at least we still get a sense that the letters not written by Peter linger on the outer edges of the cluster. Heloise's Epp. 2, 4 and 6, on the other hand, attest far less of such outlier behaviour, with exception of Ep. 2 (figs. 8.2, 8.3 and 8.3).

Therefore, in awaiting counterevidence, I stand by my interpretation. The state-of-the-art acceptance of the Heloise-Abelard collection's dual authorship is not supported by computational evidence. In fact, there is much against it. The letter collection exhibits deep affinities with Abelard's external letters, as with his sermons and even more technical works (see figs. 8.4b). By contrast, the only sample we have of Heloise's writing tends to distantiate itself from the collection (especially in fig. 8.2b). If one insists that the Heloise-Abelard collection is a genuine collection by two authors and not Abelard alone, one has to provide evidence to demonstrate —in a consistent, replicable and verifiable manner— how Heloise's letters set themselves apart from Abelard's, and certify that this method performs as desired in comparable scenarios of dual authorship (and other *epistolaria*) for the twelfth century. If Heloise contributed to the Heloise-Abelard collection at all, I could find no trace of her.

Figure 8.7: PCA plots of *epistolaria* of Peter the Venerable, Hildegard of Bingen and Heloise and Abelard. Settings are identical for both plots: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies.

## 8.6.2 *Epistolae duorum amantium*

### The Relation to the Heloise-Abelard Collection

In fig. 8.8, the *EDA* are benchmarked against the Heloise-Abelard collection, Abelard's Epp. 9–13, Heloise's letters (green star), and the *Institutiones nostrae*. As was the case in fig. 8.2, two different feature subsets were tested, represented in subplots 8.8a (MFW) and 8.8b (function words, fully listed on p. 369). Two major trends can be remarked. Firstly, in both plots, the *EDA* exhibit a style which distinguishes it from Heloise and Abelard's works. There is no reason —at least not on the current, formalistic level, which has its limitations— to believe that our <*V*>*ir* corresponds to Abelard, and our <*M*>*ulier* to Heloise.

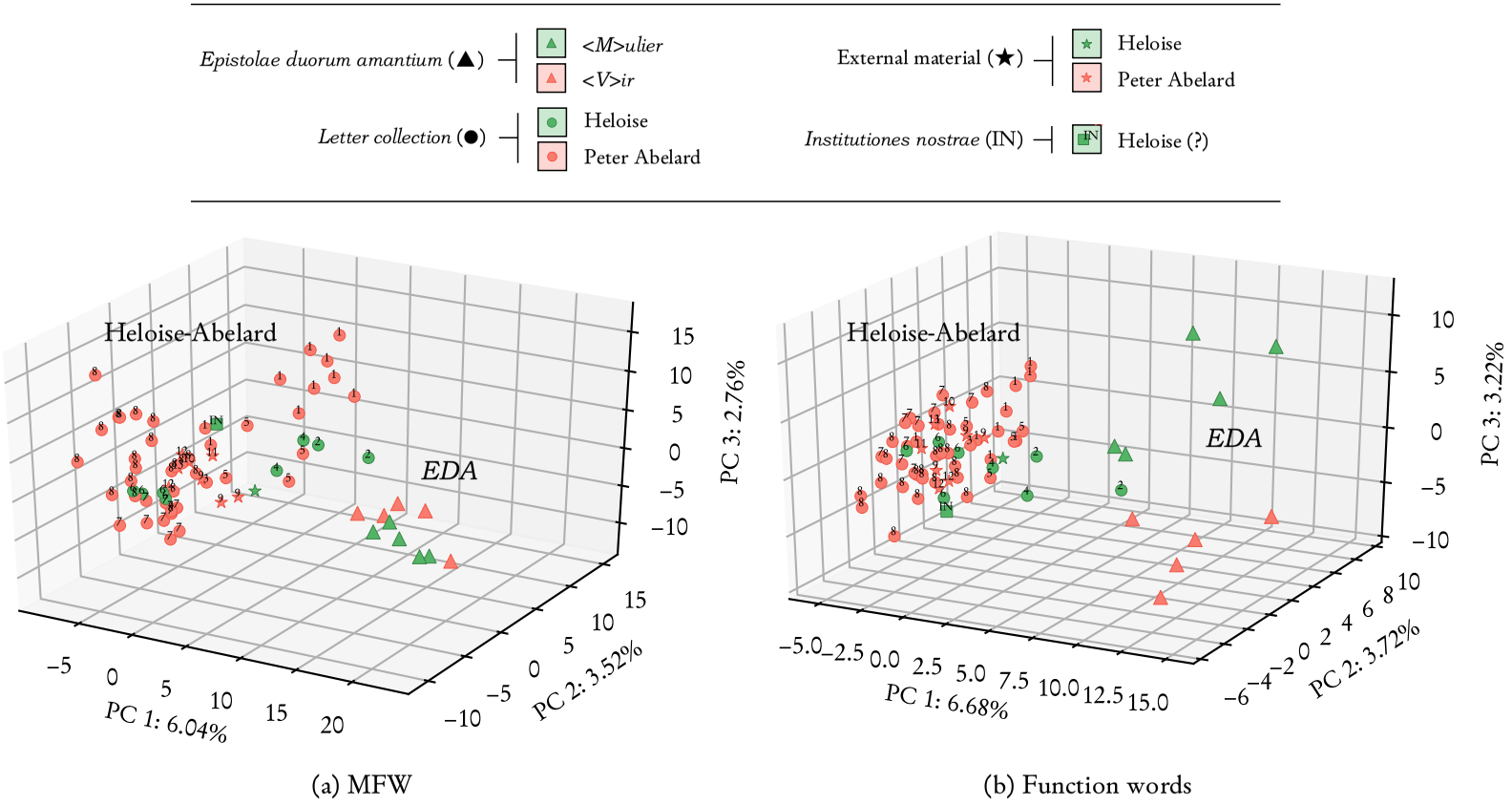
The PCA plots in fig. 8.9 demonstrate further the contrast between the Heloise-Abelard collection and the *EDA*, by individually benchmarking both letter exchanges against Abelard's canonical works (table 8.1, p. 255). As was already demonstrated earlier in subplot 8.4b, the Heloise-Abelard collection is absorbed in Abelard's dark red cluster completely. For the *EDA* no such strong stylistic affiliations to Abelard's works can be detected, which constitutes additional evidence that the collection cannot in any way be related to Heloise and/or Abelard.

### Dual Authorship

Before returning to the matter of the *EDA*'s authorship, let us review the hypothesis that two authors were responsible for its composition. There is indeed a subtle hint — more visible than in Heloise and Abelard's letter collection— that the *EDA* is of dual authorship. Especially in fig. 8.8b, the two groups of <*M*>*ulier* and <*V*>*ir* samples float in opposite directions and cluster together with samples of their own kind. This appears to confirm Orlandi and Stella's previous work (discussed earlier on p. 253), in which they reported on a distinction in the <*M*>*ulier* and <*V*>*ir*'s style —specifically their application of rhythmic cadences.

In fig. 8.10a, the subtle stylistic difference between <*M*>*ulier* and <*V*>*ir* is plotted on the basis of merely 50 function words. Indeed, internally, the *EDA* exhibit a distinction between two voices, one which we did not see occur in Heloise and Abelard's letters, for instance (compare figs. 8.2–8.3 with figs. 8.10a–b). One will note that the function words in fig. 8.10a, plotted on top of the PCA, repeat many of the lexical differences already reported by Ziolkowski in 2004, with some additions.[135] Also with 350 function words and in a three-dimensional plot (which is a list that certainly overreaches the number of total function words in the *EDA*), a difference between the male and female voice is sustained in fig. 8.10b. One can reflect here on our side-experiment with Peter the Venerable and Hildegard of Bingen's *epistolaria* earlier, where we benchmarked outgoing and incoming correspondence against each other.

[135] "Juxtaposing the use of such words brings to light striking disparities. It could be argued persuasively that the language




(a) MFW

(b) Function words

Figure 8.8: PCA plots of the *Epistolae duorum amantium* (*EDA*) in combination with the Heloise-Abelard collection. The digits overlapping the plot markers indicate where possible the letters' index (e.g. 1 = Ep. 1). Settings of plot fig. 8.8a: *s–l* = 1,100 *w* | *type* = most-frequent words | *n* = 1,000 | *vect.* = standard-scaled tfidf-weighted frequencies. Settings of plot fig. 8.8b: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies. Complete list of function words used in generating fig. 8.8b is given on p. 369.

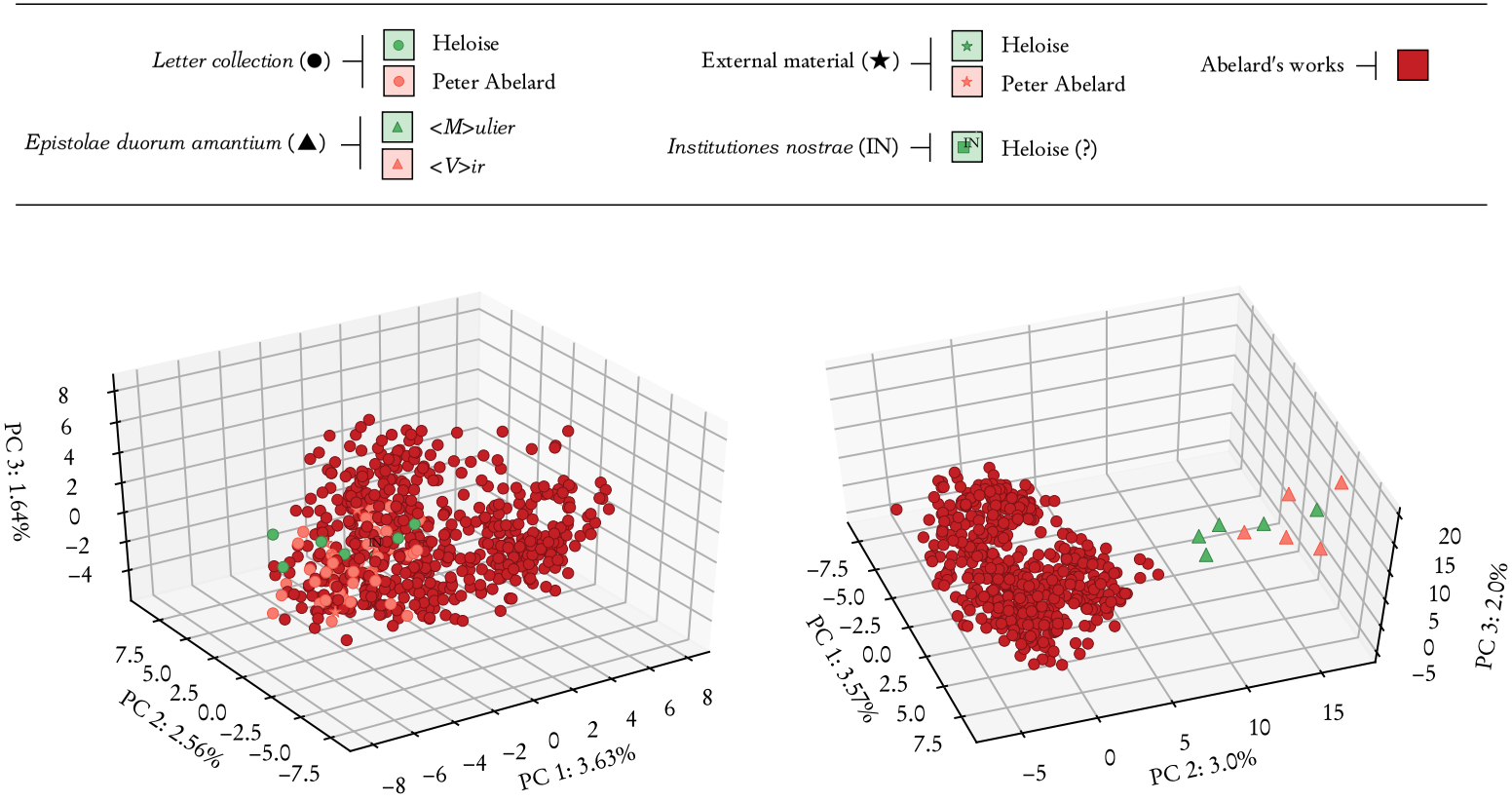



Figure 8.9: PCA plots of the Heloise-Abelard collection (fig. 8.9a) and *EDA* (fig. 8.9b) benchmarked against Abelard's canonical works. Settings of both plots: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies.

We may have found additional evidence in the *EDA* that *epistolaria* —even when composed as literary units— can contain individual contributions of style.

Note that one of two samples of Ep. 2 of the Heloise-Abelard collection seeks out the company of *EDA*. Heloise's second letter constituted an outlier for the collection as a whole earlier in figs. 8.2–8.3). This sample contains the last 1,100 words of that letter. We will return to this important point in the final paragraphs of this chapter.

## The Larger Picture: The Tegernsee Collection

Already in 1976, Dronke refrained from attributing the *EDA* to Heloise and Abelard, for the reason that "in language the Troyes collection [*EDA*] is nearest, I think, to the letters from Tegernsee; it is stylistically much further from the Abelard-Heloise collection, and from the writings of Abelard and Heloise which we know outside that collection."[136] In what follows, we involve the letters in our analysis, to test Dronke's assumption, but also to address the issues of genre and fictionality.

The Tegernsee love letters[137] are ten (not eleven)[138] short twelfth-century letters from Bavaria (totalling some 2,424 *w*), which have often been referred to for their manifest stylistic similarities to the *EDA*. Eight of them attest to have been written by women, cloistered temporarily or for life,[139] and the remaining two are by male teachers (Epp. 4 and 9). Newman contended for their genuineneness, Dronke admitted he found this precise aspect —whether the letters are public fictions or stylized, private realities— "particularly challenging to interpret."[140] The manuscript containing the letters is a large codex known as the Tegernsee manuscript, originating in

a man will employ in love letters will differ from that in his other writings, but how do we explain why one and the same man would evince such a predilection in his love letters for one conjunction (such as *quia*) over others or one Latin word for 'therefore' over another (such as the conjunction *igitur* over the particle *ergo*); why he would refrain almost entirely from a particle (*autem*) that he elsewhere favours; or why he would place an adverb (*ita*) in a position in the sentence that he otherwise regularly shunned? Finally, why should he in his love letters alone avoid completely a simple word for *certainly* for which he elsewhere displays a great fondness? These differences do not reflect modulations in formality and informality or in public and private manners of self-expression. Instead, they point to different authors with distinct ways of structuring thoughts and conveying them in words," see Ziolkowski, "Lost and Not Yet Found," 191.

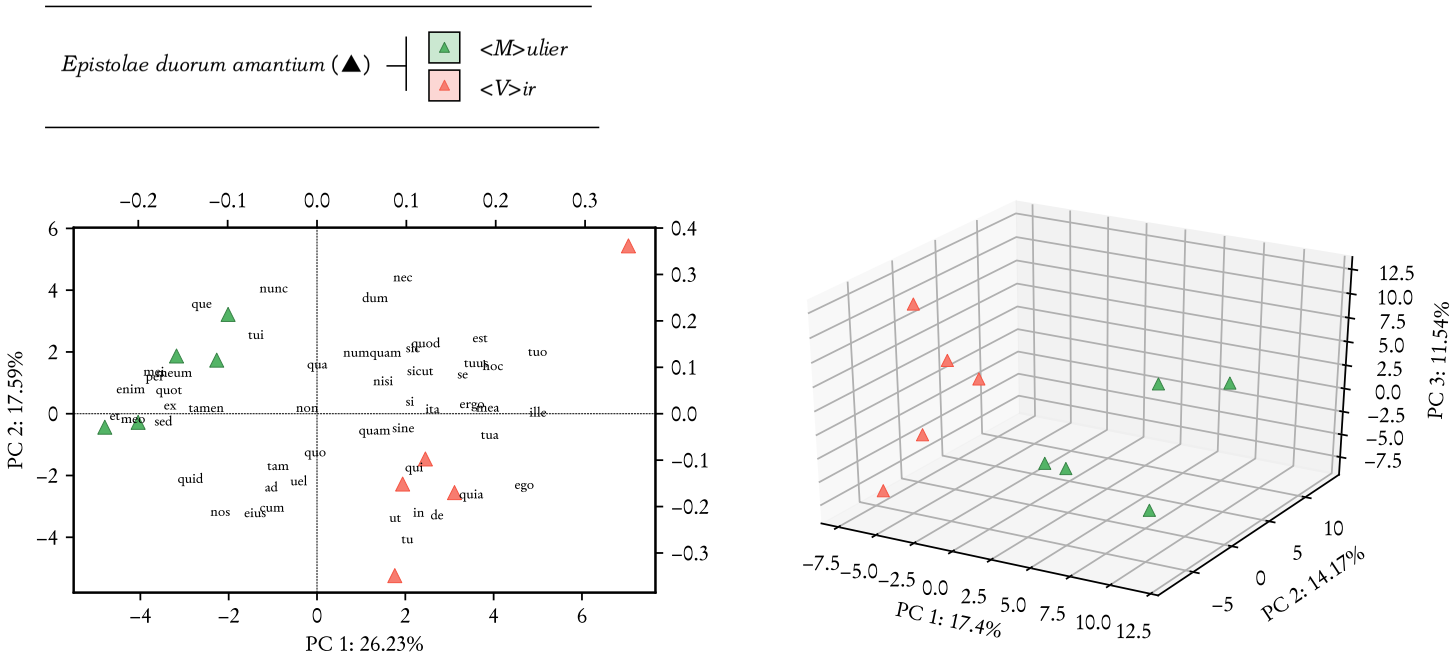
[136] Dronke, *Medieval Testimonies*, 25.

[137] Edited in Helmut Plechl and Werner Bergmann, eds., *Die Tegernseer Briefsammlung des 12. Jahrhunderts*, MGH. Epistolae 2: Die Briefe der Deutschen Kaiserzeit 2 (Hannover: Hahnsche, 2002), 343–66; and Epp. 1–7 also formerly edited and translated in Peter Dronke, *Medieval Latin and the Rise of European Love-Lyric*, 2nd ed. Vol. 2. *Medieval Latin Love-Poetry*. Oxford, UK: Clarendon Press, 1968 (1966), 472–85.

[138] Plechl and Bergmann have edited eleven letters in the *Anhang* (*Liebesbriefe*) (from 343 onward), but their Ep. 5 (with incipit "Patri suo H.") has often been left out when speaking of the Tegernsee love letters due to a lack of thematic relationship. Nevertheless, on stylistic grounds some matches have been indicated, and from a codicological point of view there may be an argument for its inclusion: the letter occurs on fo. 69$^{va}$, which is in the middle of the first series of love letters. See Plechl and Bergmann, *Die Tegernseer Briefsammlung*, xv.

[139] "But even if these love letters stem from or were sent to a convent or a foundation of canonesses, this does not necessarily mean that the writers had dedicated themselves to the religious life," see Peter Dronke, "Women's Love Letters from Tegernsee," in *Medieval Letters – Between Fiction and Document*, ed. Elisabetta Bartoli and Christian Høgel, Utrecht Studies in Medieval Literacy 33 (Turnhout: Brepols, 2015), 217.

[140] Ibid., 216.




(a) Distinction between <*V*>*ir* and <*M*>*ulier* on the basis of 50 function words.

(b) Distinction between <*V*>*ir* and <*M*>*ulier* on the basis of 350 function words.

Figure 8.10: PCA plots of the *Epistolae duorum amantium* (*EDA*) in isolation. Settings of plot fig. 8.6a: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 50 | *vect.* = standard-scaled tfidf-weighted frequencies. Settings of plot fig. 8.6b: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies.

Germany, "presumably in an abbey of nuns or canonesses,"[141] and composed before 1160–86 (although it contains materials older than that date).[142] In form and style they are clearly literary, abundant with wordplay, rhymes, parallellisms and metrical sentences (e.g. leonine hexameters). In content and theme they are often intimate, secretive and erotic, and frequently explore teacher-student relationships. Similarly as in the *EDA* (and the Heloise-Abelard collection, for that matter) we find elements of forbidden love and complaints about long absences and enduring silences.

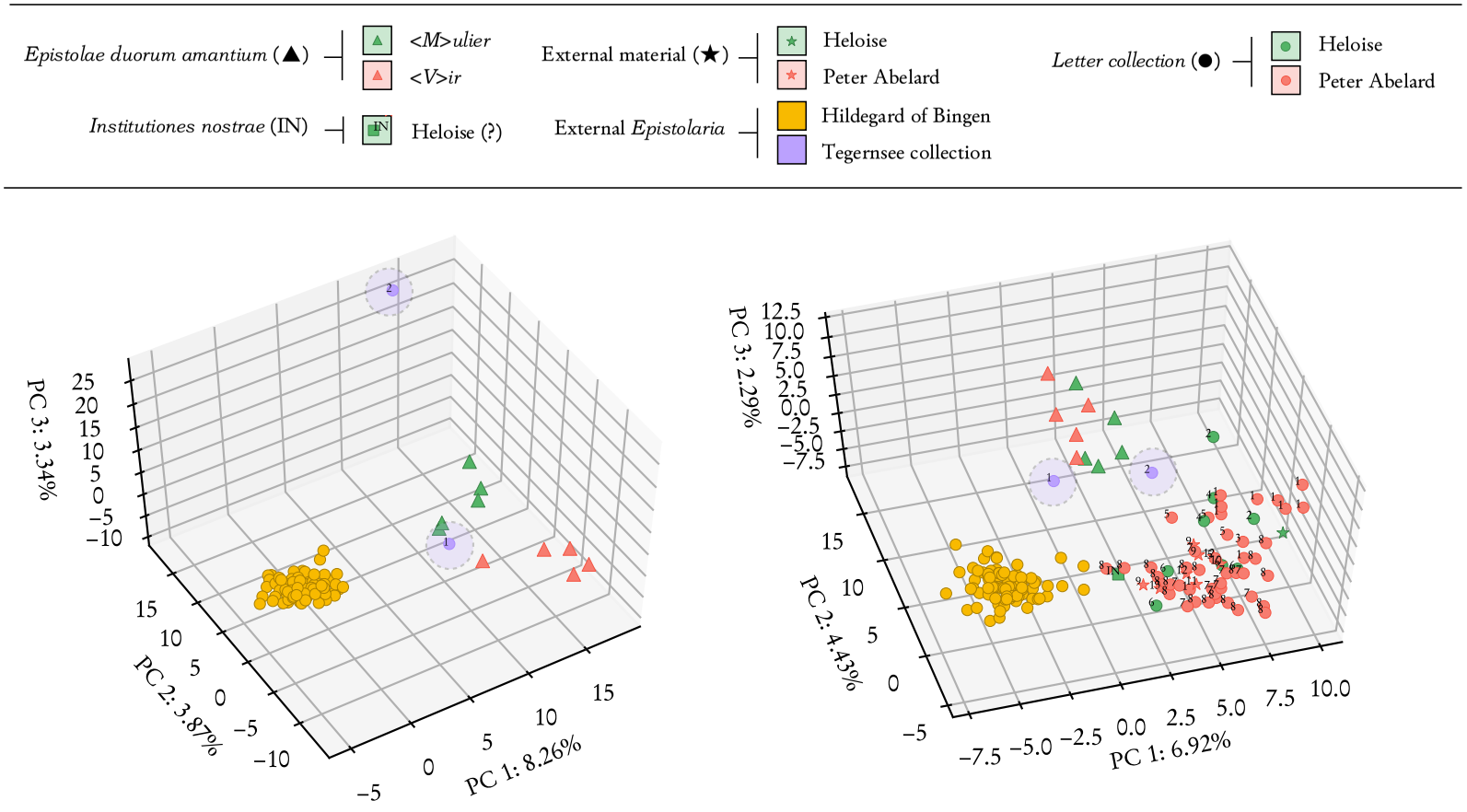
Newman, who supported the *EDA*'s ascription to Heloise and Abelard, reported and commented upon the Tegernsee collection's affinities with the *EDA*, and both texts' common sources. Nevertheless, she ruled out the possibility that both exchanges were of same authorship. Aside from the fact that these texts' transmission histories suggest that the Tegernsee collection and the *EDA* stem from different regions, such was also irreconcilable with her belief that both the *EDA* as the Tegernsee collection (allegedly written by multiple anonymous authors both male as female) are 'genuine' letters and not school exercises.

First of all, fig. 8.11 confirms Dronke's statement that the Tegernsee letters (violet-tinted samples) are stylistically far more resemblant of the *EDA* than the Heloise-Abelard collection. Especially sample 1 of the Tegernsee collection, which contains Epp. 1–7, keeps particularly close to the *EDA*. Although I do not mean to claim here that the collections' stylistic similarities are hard evidence to argue for their authorship by the same writer, I do believe one may draw conclusions in observing that the *EDA* are more akin to the Tegernsee collection than they are akin to the Heloise-Abelard collection. Such dynamics make one ponder upon the role of schooling and rhetoric as essential aspects in understanding the *EDA*. After all, would the *EDA* not, as was pointed out by Wim Verbaal in 2014,[143] perfectly match the picture of a twelfth-century school culture, the revival of anonymously transmitted Ovidian love poetry, and the romanticization of Heloise and Abelard's story already in medieval times? We may even raise further questions down this track. Could it not be disputed, in light of the letters' dissimilarity to Heloise and Abelard's style and the lack of manuscript evidence, to what extent the *EDA* are twelfth-century at all, and not form a younger *imitatio* with the objective to capture Heloise and Abelard's love story in a twelfth-century Latin style?

[141] Newman, *Making Love*, 229.

[142] The Tegernsee manuscript, containing the larger Tegernsee letter collection, corresponds to MS München, Bayerische Staatsbibliothek, Clm 19411, with the love letters contained on fos. 69$^{r}$–70$^{r}$, 100$^{v}$ and 113$^{v}$–114$^{v}$. See ibid., xv, 4–5 and 229.

[143] Verbaal, "How the West was Won by Fiction," 199.




(a) The *EDA* benchmarked against the Tegernsee collection and Hildegard of Bingen's *Epistolarium*.

(b) Rehearsal of (a), but with addition of the Heloise-Abelard collection.

Figure 8.11: PCA plots of the *EDA* and Tegernsee collection, with respectively Hildegard's *epistolarium* (fig. 8.11a) and the Heloise-Abelard collection (fig. 8.11b) as benchmark corpus. Settings of both plots: *s–l* = 1,100 *w* | *type* = most-frequent function words | *n* = 350 | *vect.* = standard-scaled tfidf-weighted frequencies.

## 8.7 Sliding *Impostors* Method

### 8.7.1 Verification

One other unsupervised method (which needs no introduction anymore) is the *impostors* method. Considering that we actually only have available specimen of Abelard's style and no reliable material for Heloise (see table 8.1 earlier), the authorship of the Heloise-Abelard collection could be approached as a one-class verification problem, where the question becomes, "Has Abelard written this specific sample of text, yes or no?" If the answer is negative, we may presume that another author was responsible for having written —or at least significantly rewritten— the text. By extension, we could also widen this question to "Is this specific sample of the same style as Abelard, <*V*>*ir*, <*M*>*ulier* or Tegernsee?" In this sense we combine the analysis of the Heloise-Abelard collection with an analysis of the *EDA*. The rationale of the *impostors* method, as we have seen above, is that the answer to that question becomes 'yes' if and only if a specific sample consistently selects the same favourite candidate $k$ out of 100 times (establishing $\sigma^*$), a candidate which competes with a number of randomly selected 'impostor' or 'distractor' documents from the benchmark corpus (p. 304).

As was the case in the chapters on Elisabeth of Schönau and Hildegard of Bingen, where we had to cope with similar issues of scant training material for all candidates, we chose for a sliding (or rolling) method. The letter collection was not divided into discrete chunks, but scanned linearly from Ep. 1 to Ep. 8. The corpus was divided into partially overlapping windows (500 $w$) and analyzed step by step (100 $w$ at a time).

Although this all sounds fine in theory, as was noted earlier on p. 119 and p. 159 in this thesis, the *impostors* method is a complex algorithm for a complex problem, and seeks to crack what can arguably be called the fundamental question of all authorship attribution.[144] No wonder it does not always succeed at what one wishes to find, and when training on Abelard's corpus, the algorithm proved deficient. At an 'optimal' $\sigma^*$ score, after attempts to train on increased and decreased numbers of candidates, different selection methods (e.g. one random document per impostor vs. random documents), across texts and with both cosine and delta distance, the results remained as follows:

Especially the precision of this model stayed alarmingly low, and, similarly as with the case of Elisabeth in chapter 5, the $\sigma^*$ hardly stabilized in the current set-up, wherefore the model was inept in expressing its confidence. This is one of those instances where one must ask if these results tell us something about Abelard's stylistic versatility, or about the model's deficiency.

[144] Koppel et al., "The "Fundamental Problem" of Authorship Attribution."

| | dev set | test set |
|---|---|---|
| **accuracy** | 0.54 | 0.55 |
| **precision** | 0.54 | 0.54 |
| **recall** | 0.67 | 0.70 |
| **f1** | 0.61 | 0.62 |

Table 8.2: Evaluation metrics for training the *impostors* method on Peter Abelard. Number of impostor documents was 20, distance metric was cosine, and documents were selected completely at random, training and testing occurred across texts.

### 8.7.2 Heloise's Second Letter

Despite my reservations on this model's worth, I have nevertheless visualized the most important trends of the *impostors* method's results in fig. 8.12 for the sake of transparency. This figure highlights all those instances where Peter Abelard was selected as favourite author for each particular 500-word sample (in red). Generally, Abelard was still selected quite frequently (65.28%), but by contrast was not always recognized in some sections of his own autobiography and letters. Is it meaningful that Abelard was not one single time selected as the favourite candidate for Ep. 2, conventionally ascribed to Heloise?

This brings us to a final point. Throughout this chapter, Heloise's Ep. 2 has occasionally come in the limelight due to its outlier behaviour, and the *impostors* method forms no exception. A peculiar result was given for this letter. Ep. 2, the most passionate of her letters, the one most Ovidian, and the one which best explores the student-teacher eroticism recurring time and time again in the *EDA* and Tegernsee collection, has a very strong match with the stylistic profile of <*V*>*ir* in the *EDA*. Not <*M*>*ulier*, but <*V*>*ir* (!).

One may rightfully be dismissive of this result, definitely in consideration of the fact that the evaluation scores earlier indicated the model's poor performance. However, a glance at the PCA plots in fig. 8.8 on p. 270 teaches us that also here the stylistic affinity between Ep. 2 and the voice of <*V*>*ir* was particularly strong. This seems to emphasize that this is no mere coincidental statistical slip. The attribution therefore gives food for thought. Since we can impossibly trace a relationship between <*V*>*ir* and Heloise's second letter exclusively, their large correspondence in terms of tone, theme and shared sources (mainly Ovid) may make one reflect on the relationship between style and tradition, and on the bearing which mimesis may have had on individual expressions of style, to the extent that a computer was misled by it some centuries later.

## 8.8 Conclusive Remarks

In a sensitive case such as that of Heloise and Abelard, which has in its history so frequently been tormented by impulsive attributions, it seems advisable not to fall for

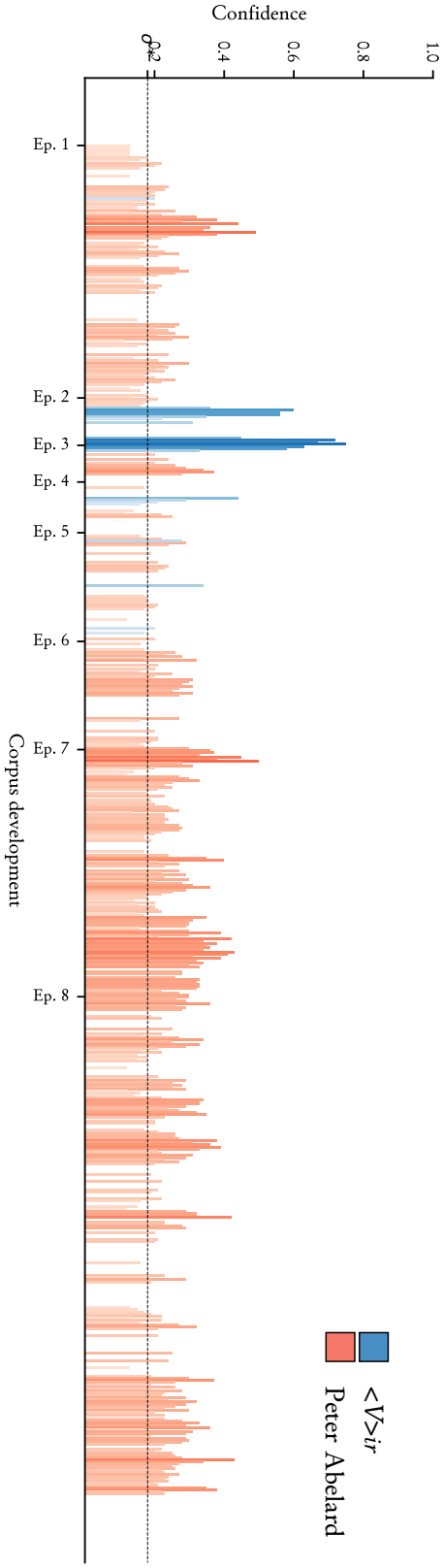



Figure 8.12: Sliding *impostors* method on the Heloise-Abelard collection. The x-axis shows the progression of the sliding windows throughout the Heloise-Abelard collection by taking a step of 100 words at a time. In total, attributions were made for 625 partially overlapping text samples consisting of 500 words. Red means Abelard, blue means <*V*>*ir*. The x-axis labels indicate where each of the letters begins. The *y*-axis (the height of the bars) and the colours' intensity indicate the confidence score of the *impostors* method. The $\sigma^*$ threshold is visualized by a dotted horizontal line. At some points Ep. 2 shows a very strong similarity with the stylistic profile of the <*V*>*ir* from the *EDA*.

the same mistake, and to refrain from all too rash conclusions. This, also, was an important point of this chapter: to show how the projection of readers' ideologies onto the past may prove misleading or even disastrous for our understanding of a period. Nevertheless, although additional research is required as to how medieval *epistolaria* were able to construct realities to the letter, I believe that two findings shed a dramatically new —or should we say, 'old'?— light on the texts that have been argued to be a result of the lovers' literary collaboration.

Firstly, scholars who have in the past denied dual authorship of the Heloise-Abelard collection by bringing to attention its stylistic unity appear to have won evidence to strengthen their claim. The general absence of a writing style attributable to Heloise is surprising to say the least, and stands in stark contrast to Hildegard's or even Elisabeth's collaborations with male writers discussed in chapters 5 and 6. It proved nearly impossible to find any trace of contribution by her. Nevertheless, the underlying reasons of how the attribution in favour of Abelard came about must remain questionable. Did Abelard invent the Heloise in these letters from scratch, or did he significantly 'rewrite' her? Is the relationship between master and pupil so exceptionally strong that Heloise's style came to stand in such close resemblance to that of Abelard's, to the extent that the lovers became inextricable? Could there be an additional reason that I have missed, one that transcends the question of authorship, for how the homogeneity of the letter collection came about? Perhaps so, although one should note that the strong and consistent stylistic connection of the letters to Abelard's canonical works excludes the hypothesis of a single forger to have composed the whole.

Secondly, there is no statistical evidence available to support the claim that Heloise and Abelard wrote the *Epistolae duorum amantium*. They may well have been inspired by the lovers' story, but the *EDA* hardly match with any stylistic pattern contained within the Heloise-Abelard collection (with exception of Ep. 2), works formerly attributed to Heloise, or works formerly attributed to Abelard. Instead, the *EDA* show a strong affiliation to the blooming love poetry of the late eleventh and twelfth century, as a comparison to the Tegernsee letters suggested. As opposed to the Heloise-Abelard connection, however, the odds of dual authorship are significantly higher for the *EDA*.

# Conclusion

Having thoroughly explored five individual case studies of twelfth-century collaborative authorship in Latin literature, we now arrive at a synthesis. Which are the factors that connect and/or distinguish the collaboratively composed writings of Bernard, Elisabeth, Hildegard, Suger, and Heloise and Abelard, and what are this thesis's main contributions to our knowledge on collaborative authorship in twelfth-century Latin literature, and the dynamics of gender, authority and synergy?

I would like to set off by focusing on three conclusive remarks of a more general nature, bringing together some of the similarities across the discussed case studies. Firstly, throughout this thesis, we have given sufficient reason for alertness when it comes to the collective aspect of medieval authorship. We have shown that the interaction of these authors with other writers has consequences for their literary accomplishment and reception. Although the evidence for collective aspects inherent to these texts might not constitute a discovery that will jump on medievalists as all too great a surprise, we have demonstrated that identifying and localizing traces of writing collaborations brings fascinating insights concerning the authors' respective historical context, their practice of composition, their style and their authority. All too often the implications this might have for the interpretation of the text easily goes lost.

A second conclusive remark which may have been, by contrast, somewhat more surprising, is that this thesis has shown that there is a case to be made for the individuality of expression in the twelfth century as well. When taking into account the fact that the composition of medieval text largely originated in mimesis with authoritative forebears and intensive participation of collaborators to the final project, it is remarkable to observe that the methodology nevertheless allows to capture individual and distinctive styles. This observation challenges some of the tenets of the New Philology, and presents an alluring trajectory for future research. Although not without its controversies, the idea that the author is a concept unfamiliar to the Middle Ages,[145] has remained an influential idea in the approach to medieval authorship in the wake of poststructuralism. That computational stylistics is often indifferent to this idea, and that medieval text reuse remains largely inconsequential for detecting individual stylistic profiles, is an insight that heeds warning. The impact of an intertextual writing

[145] "L'auteur n'est pas une idée médiévale," see Cerquiglini, *Éloge de la variante*, 25; Cerquiglini, *In Praise of the Variant*, 8.

culture, and the effect of text reuse on the DNA of authors' stylistic profiles needs further exploration.[146]

A third conclusive remark binding these case studies together is that in each of them we have encountered at least one instance where the voice of the text is contradictory to the author to have done the largest physical contribution to the text. Think of Bernard's sermons containing Nicholas's phrasing, or Elisabeth's visions on the Ursuline relics containing Ekbert's style, the Hildegardian language in Hildegard's sisters' letters on the visionary's death, Suger's impersonation of Charlemagne, and Abelard's strong appropriation —could we even say construction?— of Heloise's voice in the collection. The discrepancies between voice and physical scriptor —a term which I use here in a broader sense than 'scribe'— suggests that authorship was consciously modelled according to literary-social norms or particular agendas of the composers, envisioning a public rather than a private function. In this modelling of authorship, the notion of authority is key.

For instance, Nicholas's 'appropriation' of the sermons of Bernard of Clairvaux is to an extent justifiable, as they contain many traces of his Latin and not Bernard's. The voice and authority in the text, however, is still that of the abbot of Clairvaux, and Nicholas showcases how well he imitates an authoritative writer with whom he had been granted the honour of collaborating. Also the disjunction between Elisabeth's voice and Ekbert's dominant contribution can be explained from the viewpoint of authority. In a sensitive situation such as that of the Ursuline relics, Ekbert's own reputation was at stake, wherefore he attempted to maximize the impact of his collaboration with his sister. On the one hand he increased his stylistic grasp over the text more than elsewhere, but simultaneously he relied on Elisabeth for the words to become authoritative. Hildegard's death passage, attributed to her sisters, came in a phrasing highly reminiscent of Hildegard's own works, and revealed the visionary's exceptional authority resonating in the Rupertsberg community after she came to pass. Suger, then, relied on the authority of Charlemagne, and impersonated the legendary emperor so as to forge ties between Saint-Denis and the Capetian dynasty. Finally, the stylistic homogeneity of the Heloise-Abelard collection suggests that Abelard needed the voice of Heloise as prioress of the Paraclete so as to authorize the document. All five case studies thus demonstrated that the rhetoric of the text is constructed to make the reader lend ear to an authority which not necessarily coincides with the main physical contributor. Dismantling this disjunction between the voice in the text and its physical scriptor may present one of the most promising ways by which computational stylistics can contribute to literary theory of the Middle Ages in the future.

Then again, these three foregoing conclusive remarks should not let one resort

[146] The first initiatives in this direction are currently being undertaken for Bernard of Clairvaux. See Enrique Manjavacas, Brian Long, and Mike Kestemont, "On the Feasibility of Automated Detection of Allusive Text Reuse," in *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (Minneapolis, MN, 2019), 104–14.

to generalizations. Interesting observations have been shown not only to lurk in what binds these case studies, but precisely in what sets them apart. The historical-contextual parameters for mapping out collaborative authorship are susceptible to change from one case to the next. We have seen that the outcome of some experiments were surprising and occasionally challenged the 'traditional' categories by which writing partnerships are studied. For instance, even though cross-sex collaboration has a set of common characteristics, we have also seen that none of our three predominant female figures in this thesis strictly correspond to one another. Despite their many similarities, the results in the current thesis suggest that Hildegard of Bingen was a far more authoritative writer than Elisabeth of Schönau. Furthermore, an alarming and perhaps unanticipated contrast with expectations was the absence of Heloise, who has always been considered a celebrated highlight of female authorship in the twelfth century.

In light of these similarities and differences from one case to the next, what may consequently provide a suitable answer to the complex question of collaborative authorship in twelfth-century Latin literature? For this purpose I would like to hark back to the theoretical notion of distributional authorship as it was already mentioned in the introduction to this thesis. In conceptualizing authorship as distributional, I am not only using a description which medievalists have found to be useful for medieval authorship,[147] but I am also deliberately making associations with the normal distribution in statistics and probability theory. In what is to follow, I propose distributional authorship as a framework that is not only practical and applicable, but also possesses of the potential to reconcile the qualitative and the quantitative under a single, familiar concept.

A distribution takes into account that a reality or a phenomenon cannot be understood according to a single measure. Indeed, statistical modelling comes with an acknowledgement that any kind of empirical observation has multiple attributes and features, or, in other words, simultaneously possesses a core and yet shows deviations. In retrospect of the past chapters' explorations of twelfth-century writing collaborations, equipping ourselves with tools that allow to gauge the 'distributionality' of a phenomenon in such a way, may be beneficial to our understanding of medieval authorship, which is situated in a tension field between the individual and the collective. On the one hand, distributionality implies the inadequacy of compartmentalization —i.e. ascribing a text to single personality—, and proposes an evaluative terminology which allows for degrees, rankings, gradations and spectra —i.e., acknowledging outliers or interferences that are deviant from a normative stylistic pattern. Such a flexible and yet practical model for authorship in the twelfth century allows us to tackle the theoretical dilemma that often arises in matters of medieval authorship. At one stroke, one can witness the New Philology's variance, but also witness those in-

[147] Masters, "Distribution, Destruction and Dislocation."

stances where individual forms of expression are evidently possible. In conceptualizing authorship as distributional, we remain at once sensitive to the scattered nature of medieval authorship, as offer very precise tools for localizing individual contributions by different authors where feasible. This is where computational stylistics may make the most interesting contributions to the humanities in the future, especially if historians and literary scholars seek to determine a text's authorship, dating, and provenance, and cannot solely rely on traditional philological disciplines.

Vice versa, it must be noted that these 'traditional' disciplines still have much to offer to computational stylistics. In researching all of these authors, I have mostly been carrying out experiments from a 'comfortable' position, where the texts had been provided to me from editions. Systematic research is required as to how the material peculiarities common to medieval literature —not always taken into account in their virtual rendition— interfere with style, or have an influence on the formal properties of medieval documents. This could be done, for example, by focusing on a number of case studies which enable computational stylisticians to compare style across variant witnesses of texts. The results of such a study, paying even closer tribute to the theoretical requirements of the New Philology, could provide further insight in our understanding of collective creativity in the Middle Ages. In this thesis, the question of writing collaborations gained precedence over that of manuscript variance, but in a stylistic analysis of medieval texts the two are obviously related. It is a serious challenge, however, presenting conditions that are time-consuming and require expertise from different domains.

Despite its place in a much older line of philological history, computational stylistics is a field still in its infancy, dependent upon rapidly evolving technology. This presents a slight paradox, in the sense that this thesis constantly had to strike a balance between answering questions and questioning answers. This is no struggle unique to computational stylistics but common to most scientific methodologies, although the former is arguably far more vulnerable to growing pains than the established disciplines for analyzing historical texts. And yet, despite these methodological growing difficulties, which will quickly make some of the methods applied in this thesis to become outdated, I hope to have demonstrated that there is a timelessness to the progress computational stylistics may offer to the humanities in the future. Even if it were only to succeed in coaxing scholars out of their comfort zone, in having them question their reading methods and having them challenge the answers taken for granted, computational stylistics will have been well worth our while.

# English Summary

The collaborative aspect which typified medieval authorship, and the various ways in which writing partnerships were conducted and expressed, often poses medievalists and medieval literary scholars for a difficult challenge. Although it is well accepted in current scholarship that literary assistants were not merely instrumental in the composition of texts, but had an active and influential role in the text's final outlook and the author's remembrance for posterity, the details of these collaborations and the exact extent of influence often remain unknown. Consequently, mapping out these writing collaborations can teach us a good deal about the dynamics between the various contributors, the composition context of the text, and the text's intention. This thesis focuses specifically on monastic and/or intellectual authors writing Latin literature in the geographical region enclosing parts of France and the Holy Roman Empire in the late eleventh- and twelfth centuries.

In response to such questions on twelfth-century collaborative authorship, computational stylistics is advocated, a methodology which permits the researcher to automatically detect textual writing patterns. Not only can this method be applied as a practical means of learning more about twelfth-century authorship, it also succeeds at challenging some of the theoretical tenets in the field. In particular, it is suggested that computational stylistics may formulate an apt response to the dilemma between the medieval author as divided and subjected to the material conditions of a manuscript culture on the one hand, and the 'individual' author as (s)he is attested in the twelfth century with the 'rise of individualism' on the other. This thesis explores five case studies of twelfth-century writing partnerships in Latin, in which sensitive questions surrounding authority, synergy and gender take a prominent place. These case studies involve Bernard of Clairvaux (c. 1090-1153) and his secretary Nicholas of Montiéramey († 1176/8), Elisabeth of Schönau (1129–1164/5) and her brother Ekbert († 1184), Hildegard of Bingen (1098–1179) and her biographers, Suger of Saint-Denis (1080/1–1151) and his entourage, and the renowned twelfth-century lovers and correspondents Heloise of Argenteuil († 1164) and Peter Abelard (1079–1142). The case studies do not follow a chronological order, but are arranged in such a way that the thematic emphasis gradually shifts from the material reality of the scriptorium to the reality of the text.

**Chapter 1. Collaborative Authorship (pp. 25–53)**

Chapter 1 discusses collaborative authorship on the thematic sliding scale that runs from the scriptorium to the text. Firstly, it elaborates on the role of secretaries in the daily practice of premodern writing cultures, more especially that of the twelfth century. We posit that the influence of secretaries at once heeds warning for a larger degree of fluctuation and a decrease in authorial control. Simultaneously, we acknowledge that much depended upon individual circumstances and authors' personal authority and charisma in granting their secretaries a greater or lesser degree of liberty in the composition of their texts. The chapter discusses the literary-social norms and values which demarcated the respective spheres of influence in writing collaborations, and reflects on cross-sex collaborations and as particularly interesting case studies to explore the sensitive power dynamics at stake in medieval collaborative composition. Finally, the chapter illustrates how collaborative authorship is to be understood in a broader sense than merely by its physical and material manifestations. The twelfth-century author was also in continual 'collaboration' with a tradition and a past, especially the Latin past. In such a culture, impersonation, text re-use, anthologization and compilation are an integral part of literary composition.

**Chapter 2. Computational Stylistics (pp. 55–87)**

Chapter 2 contains a conceptual and theoretical introduction to computational stylistics, which makes it fundamentally different from chapter 3, which is of a practical/technical nature. It sets off with the discipline's historical development and a state of the art, beginning in the nineteenth century and ending with its increasing popularity today by virtue of technical advancements in the field of computer science. Amongst others, the foundational works of Mendenhall, Zipf, Yule, Mosteller and Wallace, Burrows, Pennebaker and Moretti are discussed. Consequently, this chapter asks critical questions concerning computational stylistics' often encountered claim of being revolutionary. Although this is in some regards undeniably true, this idea is critically questioned by presenting 'an alternative history' of computational stylistics, which exposes some of its concerns to run parallel with textual criticism, antiquarianism, diplomatics and stylistics. In fact, the discipline has its roots in proto-scholasticism and Abelard, and boasts an ancestry in treatises by humanists such as Alberti, Erasmus and Valla. One may also see its principles in Mabillon's diplomatics and eventually Lachmann's textual criticism in the cradle of nineteenth-century philology. After a state of the art for computational authorship attribution for Latin specifically, the chapter addresses the often encountered objectivity claim. Scholars such as Rudman have voiced fierce criticism toward the claim by pointing out the abundance and variety of stylistic features and methods proposed.

**Chapter 3. Capturing a Stylistic Profile (pp. 89–121)**

Chapter 3 offers a hands-on introduction to the common misperceptions and problems of the stylometric method, and offers a preliminary intuition of the eleventh- and twelfth-century context in which part 2's case studies are framed. Through inductive experimentation on contemporary authors such as Rupert of Deutz (c. 1075–1130), Honorius Augustodunensis (c. 1080–c. 1154), Peter the Venerable (1092–1156), Hugh of Saint-Victor (c. 1096–1141), Guibert of Nogent (c. 1053–1125), and many others, we introduce some of the concepts and difficulties that are proper to computational analysis of twelfth-century Latin literature. The chapter uses this benchmark corpus of twelfth-century authors (see p. 18) to explain the concepts of preprocessing, sampling (data segmentation), vectorization, feature extraction and feature selection, visualization (PCA and network analysis), machine learning, classification, evaluation and authorship verification (the *impostors* method).

**Chapter 4. Scandal in Clairvaux: Bernard and Nicholas of Clairvaux (pp. 125–146)**

Chapter 4, focusing on the collaboration between abbot Bernard of Clairvaux (c. 1090-1153) and his flamboyant head secretary Nicholas of Montiéramey († 1176/78), puts the tension of appropriation and authority between author and secretary firmly on the agenda. Nicholas was one of those professional, well-educated scribes siding Bernard, schooled in rhetoric and epistolography, and the headman of the scriptorium having secretaries of his own. He is known to have fallen out with Bernard after having sent out letters without the latter's consent, which lead to his expulsion from the monastery. After this event we also find Nicholas 'plagiarizing' Bernard's works, an act for which he is still often portrayed by historians and literary scholars as the Judas by Bernard's side. By diachronically assessing Nicholas's stylistic contributions in the epistolary corpus of his abbot and on some of his sermons, we challenge this commonly encountered portrayal, and assess the extent of Nicholas's stylistic contribution in order to determine the degree by which his claim onto these works can be considered justified. It turns out that not only the authority of Bernard but also the authority of modern editors, such as Leclercq and Rochais, can have a detrimental impact, as they have studied the case of Nicholas's 'text theft' through an unbefitting, anachronistic lens, thereby projecting current-day norms onto a medieval writing culture. The stylometric analysis in this chapter indicates that the majority of Bernard's sermons which Nicholas claimed as his own actually do contain his style. The chapter closes with reflections on the historical understanding of Nicholas's position as a secretary, and on Bernard's twelfth-century authorship as 'distributional' rather than individual.

**Chapter 5. Siblings' Synergy: Elisabeth and Ekbert of Schönau (pp. 147–176)**

The Benedictine nun and visionary Elisabeth of Schönau (1129–1164/5) would have uttered her visions in Latin, in German, or in a mixture of both. In redacting the Latin text to parchment, she asserted to rely on her brother Ekbert. The resulting corpus of visions constitutes a patchwork of different genres (epistolography, hagiography, sermons, visionary treatises ...), has no fixed structure, and has been transmitted in various guises. In light of all this polymorphism, the question has often been raised to what extent Ekbert's supervision has suppressed Elisabeth's presence in the text. This chapter presents results to indicate that Ekbert's stylistic interference increases in those texts composed after his arrival in Schönau in 1155, whereas Elisabeth's earlier texts exhibit a larger frequency of stylistic patterns that are decidedly different. Arguments are given on a statistical basis to assume that these 'other' stylistic signals may in fact be Elisabeth's. Two specific passages in the visionary corpus are singled out for closer study. The first of these is the ending of the third visionary book, which includes an invective against the Cathars. Even though Ekbert is announced as author of the passage, and even though scholars have asserted to have identified this presence on a stylistic plain, this intuition is completely countered in the stylometric results. This raises the necessary question if the interplay of roles as suggested by the text itself should be taken for granted, and if the logical flow of the narrative in which the speaking subject shifts from Elisabeth to Ekbert coincides with the roles taken up in the material and actual setting of their collaboration. If not, then one can suspect that Ekbert's authority and authorship is conjured up with a specific function and purpose, and that explicit, text-internal ascription does not necessarily reflect the historical circumstances of composition. Secondly, Elisabeth's visions on the unearthings of the Ursuline virgins testify of an enormous stylistic contribution by Ekbert, to the extent that any possible trace of other authorship is lacking. That Ekbert's influence is so strong in this specific vision, which was Elisabeth's most popular in the centuries to come, might teach us a thing or two about the shared interests between Schönau and the ecclesial notables from Cologne, who wished to see the Ursuline unearthings endorsed through Elisabeth's visionary authority.

**Chapter 6. Larger than Life? The *Vita Hildegardis* (pp. 177–201)**

Staying on a similar track as in the previous chapter, chapter 6 explores the case of Elisabeth of Schönau's contemporary and more experienced paragon, the visionary Hildegard of Bingen (1098–1179). In recent decades, the surfacing evidence that secretaries such as Volmar of Disibodenberg († 1173) and Guibert of Gembloux (1124–1214) might have had a hand not only in the ultimate form of her texts, but also in her image for posterity, has increasingly been gaining traction. Volmar, for instance, is known to have forged the opening letter of Hildegard's *Epistolarium*, allegedly written by Pope Eugene III, whereas Guibert played a vital role in the end redaction of the Ries-

encodex, a compilation of Hildegard's *opera omnia* gradually composed during the final years of her life. Such indications suggest that the corpus of Hildegard's texts is the result of a meticulously crafted collaborative project aiming to benefit not only the memory of the founder of a community but also the entire community she left behind. In this chapter, the collaborative authorship of the *Vita Hildegardis*, Hildegard's (auto-?)biography posthumously appended to the Riesencodex, is analyzed. The *Vita* was possibly designed to pursue her canonization. Both Hildegard and her biographers gradually contributed to the text, first in the course of the last years of her life by Hildegard and (presumably) Volmar, and later completed in the mid-1180s by end redactor Theoderic of Echternach. In between these *termini a quo* and *ante quem* the work was allegedly taken up but left unfinished by secretaries Godfrey of Disibodenberg and Guibert of Gembloux.

In light of the fact that the *Vita* is an indispensable source in gaining historical knowledge on Hildegard's life, the question has often been raised if the "Life of Hildegard" is not by dint of contributions by multiple stakeholders a larger-than-life depiction of the visionary's life course. Specifically the 'autobiographical' passages throughout the *Vita*, in which Hildegard is allegedly cited directly and is taken to recount biographical information in the first-person singular, have been approached with suspicion. The results in this chapter indicate, however, that there is no reason to presume on a statistical basis that Hildegard's language was tampered with by her biographers, and that the autobiographical passages are therefore genuine. It may even be the case, as a stylistic analysis of two short letters appended to the *Vita*'s third and final book seems to suggest, that Hildegard either passively or actively contributed to parts of the *Vita* which are not traditionally ascribed to her. Since these letters recount Hildegard's miracular deeds and her veneration after death (!), this must remain a speculative case to show how thin the line is between what is to be considered 'Hildegard' from what is to be considered 'Hildegardian.' The chapter concludes with a number of reflections on Guibert of Gembloux's potential influence in those passages traditionally ascribed to Theoderic of Echternach.

**Chapter 7. Forging Ties: Suger and the Donation of Charlemagne (pp. 203–234)**

The abbey of Saint-Denis was notorious for the number of forgeries it produced in the twelfth century. Chapter 7 assesses the authorship of one of them: a disputed but well-known Latin charter in Saint-Denis's *chartrier* that is often found titled as the 'donation of Charlemagne.' Although the diploma purports to be ninth-century by content, scholarly consensus dictates that the text is most certainly a forgery of later date. The charter depicts emperor Charlemagne granting the church abbey of Saint-Denis the privileged position of "head of all the kingdom's churches" (*caput omnium ecclesiarum regni nostri*) in the presence of a general council of leading clergy and nobility. Since it appears to place high stakes in authorizing Saint-Denis as royal

abbey with close relations to the French monarchy, it has often been presumed that the donation charter was forged in the chancery of Saint-Denis either during the abbacy of Suger of Saint-Denis or that of his secretary and later successor, abbot Odo of Deuil, and was literally employed to 'forge ties' with the Capetians who strongly identified with a Carolingian past.

However, the attribution of the charter is pestered by numerous ambiguities. No medieval manuscript nor text-internal reference allows to assign a date, author or intention to the peculiar charter. Not only Suger and Odo, but also William of Saint-Denis, Suger's secretary, and even the text's first editor, the seventeenth-century monk Jacques Doublet, have been proposed as potential forgers. Textual criticism of the contents has brought no further conclusive evidence. Some have argued that the whole charter cannot be taken seriously, and that it merely presents a literary hoax or school exercise, whereas others have emphasized that the political implications of the historical relations between Saint-Denis and the French monarchy as celebrated in the charter may have had potentially harmful or deceitful purposes. Chapter 7 not only provides evidence by which the charter may be attributed to its author, which we believe can only be abbot Suger of Saint-Denis, but also takes the attribution issue as concrete illustration to show how subjectivity and ideologically charged motives strongly colorize whichever authorship hypothesis has appeared more likely to scholars in the past. It also illustrates how the categories of forgery and plagiarism as they are known today are impossible to apply to a medieval textual culture.

**Chapter 8. Love to the Letter: Heloise and Abelard (pp. 235–279)**

Current scholarship hypothesizes —and disagreement is still ongoing— that the renowned twelfth-century lovers Heloise of Argenteuil († 1164) and Peter Abelard (1079–1142) may be held responsible for having collaboratively composed two letter collections. The first is their conventionally 'accepted' letter collection (the Heloise-Abelard collection) consisting of eight letters, which constitutes a dialogue from which the pair recounts their early life and love story and gradually develops into a foundational document for the Paraclete, the monastery founded by Abelard in 1122 and seven years later presided over by Heloise as prioress. The second collection is a set of 113 short letters exchanged between two anonymous lovers (<*V*>*ir* and <*M*>*ulier*), known as the *Epistolae duorum amantium* (*EDA*). Due to a dire lack of historical information and material sources, both of these collections have been subjected to fierce scholarly debates when it comes to their provenance, dating, intention and authorship. In the past few decades, especially the restitution of Heloise's female authorship has been a major factor in the collections' authenticity disputes.

The stylometric results in this chapter, however, suggest a strong stylistic homogeneity of the Heloise-Abelard collection in favour of Abelard, which contradicts the commonly held scholarly consensus that both Heloise and Abelard equally con-

tributed to its composition. The collection's constituent letters appear inextricable in terms of writing patterns, and finding a trace that might lead to a distinction in favour of Heloise proved impossible in the current set-up. Secondly, the *EDA* cannot be matched with the stylistic profiles of the Heloise-Abelard collection. As opposed to the latter *epistolarium*, it does, however, betray behaviour that is attributable to dual authorship, meaning that a <*V*>*ir* and a <*M*>*ulier* were responsible for composing the constituent letters of the exchange traditionally ascribed to them. Instead of resembling writings by Heloise and Abelard, the *EDA* exhibit stylistic affiliations with the Tegernsee letters, a collection of ten short twelfth-century erotic love letters originally composed in current-day Germany. The chapter concludes with general remarks on the constructed nature of medieval *epistolaria*, which often served a public rather than a private function in the Middle Ages, and were composed, revised and reconstructed with an underlying agenda.

**Conclusions**

On a first level, the thesis concludes with three remarks of a general nature. The first is that each of the case studies emphasize how alertness is due for the collective aspect of medieval authorship. The second remark is that, despite these traces of collectivity testifying of a shared textual culture, we have also seen that the methodology allows to capture individual and distinctive styles nevertheless, making individuality and self-expression a possibility in the twelfth century. Thirdly, each of these case studies have shown instances where the voice of the text is contradictory to the author to have done the largest physical contribution to the text. This suggests that authorship was consciously modelled according to literary-social norms or particular agendas of the composers, envisioning a public rather than a private function. On a second level, interesting observations also lie in what sets these case studies apart: not two of the explored partnerships were strictly identical. The exposed variability of twelfth-century writing partnerships indicates the problems inherent to theoretical frameworks of medieval authorship, which often tend to make generalizations. As a possible response to this, the dissertation concludes by proposing the notion of 'distributional authorship.' A distribution takes into account that a reality or a phenomenon cannot be understood according to a single measure, but that any kind of empirical observation has multiple attributes and features.

# Dutch Summary

Het collaboratieve aspect dat middeleeuws auteurschap typeerde, en de verschillende wijzen waarop zulke samenwerkingen gevoerd werden en tot uitdrukking kwamen, plaatst onderzoekers van de middeleeuwen dikwijls voor een uitdaging. Hoewel de huidige stand van onderzoekt erkent dat schrijfmedewerkers niet enkel een instrumentele rol vervulden in het componeren van teksten, maar een actieve en invloedrijke positie opnamen in het uiteindelijke ontwerp van de tekst en de gedachtenis van de auteur voor diens nakomelingen, blijven de details van van deze samenwerkingen en de precieze mate van hun invloed vaak onduidelijk. Bijgevolg zou het in kaart brengen van deze samenwerkingen ons heel wat kunnen leren over de dynamiek tussen de verschillende bijdragers, de context waarin de tekst werd opgesteld, en de intentie van de tekst. Deze thesis focust zich specifiek op monastieke en/of intellectuele auteurs die in het Latijn schreven, gelegen in delen van Frankrijk en het Heilige Roomse Rijk in de late elfde en twaalfde eeuw.

Als antwoord op de vraag van collaboratief auteurschap wordt computationele stilistiek opgeworpen, een methodologie die de onderzoeker in staat stelt schrijfpatronen automatisch te detecteren. Niet enkel kan de methode zo een praktisch instrument aanbieden om meer te leren over twaalfde-eeuws auteurschap, ze slaagt er ook in enkele van de theoretische principes in het veld te bevragen. Zo wordt geopperd dat de computationele stilistiek een adequaat antwoord kan bieden op het dilemma tussen de middeleeuwse auteur als 'verdeeld' en onderworpen aan de materiële voorwaarden van een handschriftcultuur aan de ene hand, en de 'individuele' auteur zoals hij of zij steeds vaker geattesteerd wordt in de twaalfde eeuw met de geleidelijke 'opkomst van het individu.' Deze thesis onderzoekt vijf gevalstudies van twaalfde-eeuwse schrijfsamenwerkingen in het Latijn, waarin gevoelige vragen omtrent autoriteit, synergie en gender een prominente plaats innemen. In deze vijf casussen gaat het respectievelijk om Bernardus van Clairvaux (c. 1090-1153) en zijn secretaris Nicolas van Montiéramey († 1176/8), Elisabeth van Schönau (1129–1164/5) en haar broer Ekbert († 1184), Hildegard van Bingen (1098–1179) en haar biografen, Suger van Saint-Denis (1080/1–1151) en zijn entourage, en de bekende twaalfde-eeuwse geliefden en correspondenten Héloïse van Argenteuil († 1164) en Petrus Abelardus (1079–1142). De gevalstudies volgen geen chronologische volgorde, maar zijn op zulke manier gerangschikt dat het thematische accent gradueel verschuift van de materiële realiteit in het

scriptorium naar de realiteit van de tekst.

**Hoofdstuk 1. Collaboratief Auteurschap (pp. 25–53)**

Hoofdstuk 1 bespreekt collaboratief auteurschap vanuit de thematische progressie van de thesis, die van het niveau van het scriptorium naar het niveau van de tekst opstijgt. Ten eerste behandelt het hoofdstuk de rol van secretarissen in de dagelijkse praktijken van de middeleeuwste schriftcultuur, specifiek die van de twaalfde eeuw. We formuleren de stelling dat de invloed van secretarissen waakzaamheid vergt wat betreft de grote mate waarin de vorm van de tekst fluctueert en waarop de auteur zijn of haar controle over de tekst verliest. Tegelijkertijd stellen we vast dat op deze punten veel afhankelijk is van particuliere omstandigheden, van de persoonlijke autoriteit en het charisma van de auteur in kwestie, en van de mate waarop hij of zij de invloed van anderen tolereerde in de literaire compositie. Dit hoofdstuk bespreekt de literair-sociale normen en waarden die de respectievelijke invloedssferen afbakenden in schriftelijke samenwerkingen, en staat geruime tijd stil bij samenwerkingsverbanden tussen de geslachten. De verhoudingen tussen man en vrouw in middeleeuwse schriftelijke samenwerkingen zijn bijzonder interessante casussen om de gevoeligheid van dit type machtsdynamieken bloot te leggen. Tenslotte illustreert dit hoofdstuk dat men onder collaboratief auteurschap meer kan verstaan dan louter samenwerking in haar fysieke en materiële verschijningsvorm. De twaalfde-eeuwse auteur beriep zich voortdurend op een samenwerking met een literaire traditie en een verleden dat opgetekend was in het Latijn. In een dergelijke cultuur worden intertekstualiteit, impersonatie, het aanleggen van anthologieën en compilatie een integraal onderdeel van elke literaire onderneming.

**Hoofdstuk 2. Computationale stilistiek (pp. 55–87)**

Hoofdstuk 2 bevat een conceptuele en theoretische inleiding tot de computationele stilistiek, hetgeen dit hoofdstuk fundamenteel onderscheidt van hoofdstuk 3, dat van een meer praktisch/technische aard is. Het hoofdstuk vangt aan met het schetsen van de historische ontwikkeling van de discipline en een wetenschappelijke stand van zaken, die begint in de negentiende eeuw en eindigt met de hedendaagse popularisering van het veld dankzij de technologische ontwikkelingen in de computerwetenschappen. Naast vele anderen, worden de fundamentele werken van Mendenhall, Zipf, Yule, Mosteller en Wallace, Burrows, Pennebaker en Moretti besproken. Vervolgens stelt het hoofdstuk kritische vragen naar de objectiviteitsbewering die vaak aangetroffen wordt in het veld, en niet zelden als revolutionair wordt aangedragen. Ook al valt dit op vele vlakken te verdedigen, wordt dit idee ook kritisch bevraagd, door een ‘alternatieve geschiedenis’ van computationele stilistiek uit te tekenen. Hieruit wordt duidelijk dat het enkele van haar onderzoeksprincipes deelt met tekstkritiek, antiquarianisme, diplomatiek en traditionele stilistiek. In feite kan de discipline zelfs teruggevoerd worden op de

proto-scholastiek en Abelardus, en heeft zij haar voorouders in de traktaten van humanisten zoals Alberti, Erasmus en Valla. Men kan dezelfde principes reeds herkennen in Mabillons diplomatiek en Lachmanns tekstkritiek die de negentiende-eeuwse filologie op de kaart zette. Na een stand van zaken voor computationale stilistiek voor het Latijn, spitst het hoofdstuk zich toe op de objectiviteitsbewering. Wetenschappers zoals Rudman hebben hier vaak uitvoerige kritiek op geformuleerd wegens de overvloed en variëteit aan gehanteerde stijlherkenningscriteria en methoden.

**Hoofdstuk 3. Het opspeuren van stijlprofielen (pp. 89–121)**

Hoofdstuk 3 biedt een praktische inleiding tot de gebruikelijke misverstanden en problemen die zich aandienen bij de stilometrische methode, en geeft bovendien een eerste intuïtie van de elfde- en twaalfde-eeuwse context waarin de gevalstudies van deel 2 dienen te worden geplaatst. Door middel van inductieve proefnemingen met contemporaine auteurs zoals Rupert van Deutz (c. 1075–1130), Honorius Augustodunensis (c. 1080–c. 1154), Petrus Venerabilis (1092–1156), Hugo van Saint-Victor (c. 1096–1141), Guibert van Nogent (c. 1053–1125), en vele anderen, introduceren we enkele van de concepten en moeilijkheden die eigen zijn aan de computationele analyse van twaalfde-eeuwse Latijnse literatuur. Het hoofdstuk gebruikt dit referentiecorpus van twaalfde-eeuwse auteurs (zie p. 18) om de concepten van preprocessors, tekstsplitsing (segmentering van data), het voorstellen van tekst in vectoren, het extraheren of selecteren van stijleigenschappen (of *features*), visualisering (PCA en netwerkanalyse), machinaal leren, classificatie, evaluatie en auteurschapsverificatie (de *impostors method*) te verduidelijken.

**Hoofdstuk 4. Schandaal in Clairvaux: Bernardus en Nicolas van Clairvaux (pp. 125–146)**

Hoofdstuk 4, dat zich richt op de samenwerking tussen abt Bernardus van Clairvaux (c. 1090-1153) en dienst flamboyante hoofdsecretaris Nicolas van Montiéramey († 1176/78), plaatst de spanning tussen toe-eigening en autoriteit tussen auteur en secretaris bovenaan de agenda. Nicolas was een van de professionele, welopgeleide secretarissen aan de zijde van Bernardus. Hij was geschoold in retoriek en briefliteratuur, en een hoofdfiguur in het scriptorium die bovendien zelf over secretarissen kon beschikken. Nicolas viel in ongenade bij Bernardus nadat hij brieven had uitgestuurd zonder diens toestemming, hetgeen leidde tot zijn verbanning uit het klooster. Enige tijd later in de geschiedenis treffen we Nicolas opnieuw aan, en lijkt hij Bernardus' werken te 'plagiëren', een daad die hem in de latere historiografie de reputatie van de Judas aan Bernardus' zijde bezorgde. Door de diachrone, stilistische bijdragen van Nicolas' stijl in Bernardus' brievencorpus en sommige van diens preken te analyseren, wordt dit traditionele beeld van Nicolas in vraag gesteld, en tonen we aan dat Nicolas' stilistische bijdrage groot genoeg was om zijn toe-eigening te legitimeren. Zo blijkt dat niet enkel

de autoriteit van Bernardus zelf maar ook die van moderne uitgevers, zoals Leclercq en Rochais, schadelijk kunnen zijn, omdat ze zich in de studie van Nicolas' 'tekstdiefstal' beroepen op ongepaste, anachronistische projecties van hedendaagse normen op de middeleeuwse schriftcultuur. De stilometrische analyse in dit hoofdstuk geeft aan dat de meerderheid van Bernardus' preken die Nicolas als eigen werk opeist effectief kenmerken van diens stijl vertonen. Het hoofdstuk eindigt met beschouwingen over het historisch begrip van Nicolas' positie als secretaris in Clairvaux, en over Bernardus' twaalfde-eeuwse auteurschap als 'distributioneel' in plaats van 'individueel.'

**Hoofdstuk 5. Synergie van zus en broer: Elisabeth en Ekbert van Schönau (pp. 147–176)**

De benedictijnse non en zieneres Elisabeth van Schönau (1129–1164/5) zou haar visioenen in het Latijn, in het Duits, of in een combinatie van beide hebben uitgesproken. In het redigeren van de Latijnse tekst op perkament deed ze beroep op haar broer Ekbert. Het resulterende corpus van haar visioenen is een lappendeken van verschillende genres (briefliteratuur, hagiografie, sermoenen, visionaire traktaten ...), heeft geen vaste structuur, en werd in verschillende verschijningsvormen opgetekend. Met het oog op dit polymorfisme is in het verleden reeds vaak de vraag geopperd in welke mate Ekberts supervisie over de eindredactie ervoor gezorgd heeft dat Elisabeth's aanwezigheid in de tekst werd onderdrukt. Dit hoofdstuk presenteert enkele resultaten om aan te tonen dat Ekberts stilistische revisie toenam in de teksten die Elisabeth componeerde na haar broers aankomst in Schönau in 1155. De vroegere teksten, daarentegen, vertonen stilistische patronen die niet aan Ekbert toe te kennen zijn. Op statistische basis worden er enkele argumenten aangedragen om te suggereren dat deze 'andere' stilistische signalen van Elisabeth kunnen zijn. Verder worden er twee specifieke passages in het corpus uitgelicht voor detailanalyse. De eerste is het einde van het derde visionaire boek, dat een invectief tegen de Katharen bevat. Ook al wordt Ekbert in de tekst zelf aangekondigd als auteur van de passage, en ook al hebben onderzoekers vaak beweerd diens stijl hier te herkennen, spreken de stilometrische resultaten deze intuïtie tegen. Dit doet noodzakelijkerwijze de vraag rijzen of de wisselwerking tussen de auteursrollen zoals gesuggereerd in de tekst zelf als vanzelfsprekend moeten aangenomen worden, en of de logische progressie van het narratief waarin de spreker verschuift van Elisabeth naar Ekbert noodzakelijk samenvalt met de rollen zoals ze opgenomen werden in de materiële en eigenlijke omstandigheden van de literaire samenwerking. Ten tweede vertonen Elisabeths visioenen over de opgraving van Ursula's elfduizend maagden een enorme stilistische bijdrage van Ekbert, tot die mate dat enig spoor van andere bijdrage vrijwel afwezig is. Dat de invloed van Ekbert zo sterk blijkt in uitdrukkelijk dat visioen dat Elisabeths meest populaire werk zou worden in de komende eeuwen, leert ons het een en ander over de wederzijdse belangen tussen Schönau en de kerkelijke autoriteiten in Keulen, die de opgravingen van de mythische

maagden wensten te authenticeren met behulp van Elisabeths profetische gave.

**Hoofdstuk 6. Groter dan het leven: de *Vita Hildegardis* (pp. 177–201)**

In hoofdstuk 6 blijven we op een gelijkaardig pad als in het vorige hoofdstuk, en verkennen we Elisabeths contemporaine en meer ervaren voorbeeld, de profetes Hildegard van Bingen (1098–1179). In de afgelopen jaren werd er in aanzienlijke mate bewijs aangevoerd om aan te tonen dat secretarissen zoals Volmar van Disibodenberg († 1173) en Guibertus van Gembloers (1124–1214) niet enkel bijdroegen aan de uiteindelijke vorm van Hildegards teksten, maar ook aan het beeld dat zij moest uitdragen voor haar nakomelingen. Van Volmar weten we bijvoorbeeld dat hij de openingsbrief van Hildegard's *Epistolarium*, dat onder het auteurschap van paus Eugenius III werd gepresenteerd, vervalste. Guibertus speelde op zijn beurt een beduidende rol in de eindredactie van de Riesencodex, een compilatie van Hildegards *opera omnia* die tijdens de laatste jaren van haar leven gradueel werd samengesteld. Dit geeft aan dat het corpus van Hildegards teksten het resultaat is van een meticuleus geconstrueerd samenwerkingsproject dat niet alleen de herinnering va n de stichter van een gemeenschap in goede zin wenste te beïnvloeden maar ook die van de gemeenschap die ze achterliet. In dit hoofdstuk wordt het collaboratief auteurschap van de *Vita Hildegardis* geanalyseerd, Hildegards (auto-?)biografie die na haar dood aan de Riesencodex werd aangehecht.

Aangezien de *Vita* een buitengewoon belangrijke bron is voor de historische kennis over Hildegards leven, werd reeds vaak de vraag geopperd of het "leven van Hildegard" niet dankzij de aanvullingen van meerdere belanghebbenden een sterk geaugmenteerde aftekening van het leven van de zieneres is. Vooral de 'autobiografische' passages in de *Vita*, waarin Hildegard direct geciteerd wordt en haar lezers biografische informatie in de eerste persoon enkelvoud mededeelt, werden reeds regelmatig met skepsis benaderd. De resultaten in dit hoofdstuk geven echter aan dat er geen statistisch gegronde reden is om aan te nemen dat Hildegards taal werd aangepast door haar biografen, en dat het autobiografische materiaal bijgevolg authentiek blijkt. Het zou zelfs kunnen, zoals een stilistische analyse van twee korte brieven aan het einde van het derde en laatste boek van de *Vita* doet vermoeden, dat Hildegard heeft bijgedragen (zij het passief of actief) aan delen van de biografie die haar traditioneel niet worden toegeschreven. Aangezien deze brieven rapporteren over Hildegard's miraculaire daden en haar verering na haar dood (!), is dit een speculatieve bevinding die aantoont hoe dun de lijn is tussen een stijl die toeschrijfbaar is aan 'Hildegard' of een stijl die 'Hildegardiaans' is.

**Hoofdstuk 7. Banden smeden: Suger en de schenkingsoorkonde van Karel de Grote (pp. 203–234)**

De abdij van Saint-Denis is berucht omwille van het aantal vervalsingen dat zij produceerde in de twaalfde eeuw. Hoofdstuk 7 onderzoekt het auteurschap van een van dergelijke vervalsingen: een betwiste maar welgekende Latijnse oorkonde in de chartrier van Saint-Denis die doorgaans de titel 'schenkingsoorkonde van Karel de Grote' toegekend krijgt. Hoewel de oorkonde zichzelf inhoudelijk voorstelt als negendeeeuws, hebben onderzoekers de consensus bereikt dat het hier ongetwijfeld om een vervalsing van latere datum gaat. De oorkonde beschrijft hoe keizer Karel de Grote de abdij van Saint-Denis de bevoorrechte positie als "hoofd van alle kerken in het koninkrijk"toekent (*caput omnium ecclesiarum regni nostri*), en dit in de aanwezigheid van een algemene raad van gezaghebbende clerus en adel. Aangezien de oorkonde erg inzet op het authoriser<en van Saint-Denis als koninklijke adbij, die dichte banden met de Franse monarchie, is men er vaak van uitgegaan dat de schenkingsoorkonde werd vervaardigd in de kanselarij van Saint-Denis onder het abbatiaat van Suger van Saint-Denis of dat van diens secretaris en latere opvolger, abt Odo van Deuil. De intentie van de oorkonde was letterlijk het 'smeden van banden' met het Huis Capet, dat zichzelf sterk identificeerde met het Karolingische verleden.

Het toekennen van de oorkonde wordt bemoeilijkt door vele dubbelzinnigheden. Er is geen middeleeuws handschrift noch een tekst-interne referentie die toelaat om de datum, auteur of intentie van deze bijzondere oorkonde te achterhalen. Niet enkel Suger en Odo, maar ook William van Saint-Denis, Sugers secretaris, en zelfs de eerste uitgever van de tekst, de zeventiende-eeuwse monnik Jacques Doublet, werden in het verleden als mogelijke vervalsers gesuggereerd. Tekstkritiek van de inhoud heeft echter geen sluitend bewijs kunnen aanleveren wat betreft auteurschap. Sommige onderzoekers hebben beargumenteerd dat de hele oorkonde niet eens serieus kan genomen worden, en dat het gaat om een literair spel of een schooloefening, terwijl anderen net nadruk hebben gelegd op het feit dat de politieke implicaties van de historische verbintenissen tussen Saint-Denis en de Franse monarchie zoals die worden gecultiveerd in de oorkonde mogelijks schadelijke of bedrieglijke intenties verraden. Hoofdstuk 7 draagt niet enkel bewijs aan dat de oorkonde aan Suger van Saint-Denis kan toegekend worden, maar grijpt deze casus ook aan als concrete illustratie van hoe subjectiviteit en ideologisch geladen motieven onderzoekers in het verleden sterk hebben gestuurd in wat voor hen de meest waarschijnlijke auteurschapshypothese lijkt. De casus illustreert ook hoe de categorieën van vervalsing en plagiaat zoals ze gehanteerd worden in de eenentwintigste eeuw onmogelijk te verzoenen zijn met een middeleeuwse schriftcultuur.

**Hoofdstuk 8. Liefde tot op de letter: Héloïse en Abelardus (pp. 235–279)**

De huidige stand van onderzoek gaat uit van de hypothese —en er is nog steeds een grote mate van onenigheid— dat het befaamde twaalfde-eeuwse koppel Héloïse van Argenteuil († 1164) en Petrus Abelardus (1079–1142) mogelijks verantwoordelijk was voor het collaboratief auteurschap van twee brievenverzamelingen. De eerste is de conventioneel 'aanvaarde' brievencollectie (de Héloïse-Abelard brievencollectie) die bestaat uit acht brieven, en een dialoog bevat van waaruit de geliefden verslag doen over hun vroege leven en hun liefdesverhaal, voordat de collectie geleidelijk aan ontwikkelt naar een legitimatie voor de Paracleet, het klooster dat werd gesticht door Abelardus in 1122 en zeven jaar later werd geleid door Héloïse als priores. De tweede verzameling bestaat uit een correspondentie van 113 korte brieven tussen twee anonieme geliefden (<*V*>*ir* en <*M*>*ulier*), die bekend staan als de *Epistolae duorum amantium* (*EDA*). Door een gebrek aan historisch bewijs en materiële bronnen zijn beide collecties onderworpen aan verhitte academische debatten wat betreft hun afkomst, datering, intentie en auteurschap. In de laatste vijftig jaar speelde voornamelijk het herstel van Héloïses vrouwelijke auteurschap een belangrijke factor in de authenticiteitskwesties van beide verzamelingen.

De stilometrische resultaten in dit hoofdstuk geven echter aan dat er een sterke, stilistische homogeneïteit spreekt vanuit de Héloïse-Abelard collectie ten voordele van Petrus Abelardus, hetgeen de conventioneel geworden academische consensus dat Héloïse en Abelardus in gelijke mate aan de verzameling hebben bijgedragen tegenspreekt. De respectievelijke briefeenheden in de collectie lijken niet te onderscheiden op basis van schrijfpatronen, en er is geen spoor dat naar Héloïses schrijfstijl of haar bijdrage kan leiden in het gehandhaafde statistische kader. Ten tweede vertonen de *EDA* geen zichtbare gelijkenis met het stilistische profiel van de Héloïse-Abelard collectie. In tegenstelling tot het laatstgenoemde epistolarium, vertonen de *EDA* dan weer wel gedrag dat toeschrijfbaar is aan dubbel auteurschap, hetgeen betekent dat er effectief een <*V*>*ir* en een <*M*>*ulier* verantwoordelijk waren voor het auteurschap van de briefeenheden die hen traditioneel toegewezen worden. In plaats van overeenkomsten te hebben met de teksten van Héloïse en Abelardus, blijken de *EDA* vooral sterke gelijkenissen te hebben met de Tegernsee brieven, een collectie van tien korte twaalfde-eeuwse erotische liefdesbrieven die origineel opgetekend werden in het huidige Duitsland. Het hoofdstuk eindigt met enkele algemene conclusies betreffende de sterk artificiële aard van middeleeuwse *epistolaria*, die vaak een openbare en geen private functie hadden in de middeleeuwen, en gecomponeerd, gereviseerd en gereconstrueerd werden met een welbepaalde agenda in het achterhoofd.

**Conclusies**

Op een eerste niveau sluit deze thesis af met drie opmerkingen van een algemene aard. Ten eerste werd duidelijk doorheen de gevalstudies dat men alert moet zijn voor het

collectieve aspect dat middeleeuws auteurschap typeert. Ten tweede moet onderkend worden dat, ondanks deze sporen van collectiviteit die getuigen van een gedeelde tekstuele cultuur, de methodologie ons ook in staat stelde om individuele en distinctieve stijlen op te sporen, hetgeen individualiteit en zelfexpressie in de twaalfde eeuw een mogelijkheid maken. Ten derde bevatten elk van de gevalstudies passages waarin de stem in de tekst niet overeenstemt met de auteur die daadwerkelijk de grootste, fysieke bijdrage heeft geleverd tot het eindresultaat. Dit suggereert dat auteurschap bewust werd gemodelleerd volgens literair-sociale normen of specifieke onderliggende agenda's van de auteurs, die bovendien vaak de publieke (en niet de private) functie van de tekst verraadt. Op een tweede niveau moet vastgesteld worden dat men ook interessante conclusies kan trekken uit wat deze gevalstudies nu net onderscheidt: geen enkele van deze samenwerkingen was identiek met een andere. Deze blootgelegde variabiliteit van twaalfde-eeuwse schrijfsamenwerkingen geeft trouwens de problemen aan die men aantreft in theoretische kaders over middeleeuws auteurschap, die vaak aanleiding geven tot veralgemeningen. Als een mogelijk antwoord op deze kwestie, sluit de dissertatie af door de notie van 'distributioneel auteurschap' naar voren te schuiven. Een distributie neemt immers in rekening dat een realiteit of een fenomeen niet kan gevat worden volgens een enkelvoudige meting, maar dat elk type van empirische observatie over meerdere attributen en eigenschappen beschikt.

# Tabula Siglorum

| | |
|---|---|
| *AASS* | *Acta Sanctorum* |
| *BHL* | *Bibliotheca Hagiographica Latina* |
| *CC CM* | *Corpus Christianorum. Continuatio Mediaevalis* |
| DH | Digital Humanities |
| Ep(p). | epistola(e) |
| *expl. var.* | explained variance |
| fl. | floruit (flourished) |
| fo(s). | folio(s) |
| *Ldl.* | *Libelli de lite* |
| MFW | most frequent words |
| LLT | Brepols Library of Latin Texts |
| *MGH* | *Monumenta Germaniae Historica* |
| MS(S) | manuscript(s) |
| *n* | number |
| NLP | Natural Language Processing |
| *PL* | J.-P. Migne, ed., *Patrologia Latina cursus completus* (Paris, 1844–64). |
| *s-l* | sample length |
| *SBO* | *Sancti Bernardi Opera* |
| *SC* | *Sources Chrétiennes* |
| *vect.* | vectorization method |
| *w* | words |
| $\mu$ | mean |

# Appendix A

# Appendix

## A.I General Addenda

Table A.I.1: Benchmark corpus of eleventh- and twelfth-century Latin texts.

| Author | Texts | Edition | *n* words |
|---|---|---|---|
| Alan of Lille (c. 1128–c. 1203) | *Anticlaudianus* | Charles de Visch, ed., *PL* 210:482–578. | 38,279 |
| *Alanus de Insulis* | *Contra haereticos* | Ibid., *PL* 210:305–428. | 54,262 |
| | *De arte praedicatoria* | Ibid., *PL* 210:109–97. | 40,320 |
| | *De planctu naturae* | Ibid., *PL* 210:439–81. | 20,725 |
| | *Elucidatio in Cantica canticorum* | Ibid., *PL* 210:51–108. | 24,504 |
| | *Sermones* | Ibid., *PL* 210:197–252. | 14,930 |
| | *Summa "Quoniam homines"* | Palémon Glorieux, ed., *Archives d'histoire doctrinale et littéraire du Moyen Âge* 20, 119–364. | 90,158 |
| | | | 293,178 |
| Anselm of Canterbury (1033–1109) | *Cur deus homo* | Gabriel Gerberon, ed., *PL* 158:360–432. | 28,583 |
| *Anselmus Cantuariensis* | *Monologion* | Ibid., *PL* 158:141–223. | 33,538 |
| | *Proslogion* | Franciscus Salesius Schmitt, ed., *Opera omnia*, 93-122. | 6,591 |
| | | | 68,712 |
| Anselm of Laon († 1117) | *Enarrationes in Apocalypsin* | Maternus Cholinus, ed., *PL* 162:1499–1586. | 46,343 |
| *Anselmus Laudunensis* | *Enarrationes in Cantica canticorum* | Ibid., *PL* 162:1187–1228. | 21,217 |
| | | | 67,560 |
| Bruno of Cologne (1030–1101) | *Expositio in epistolas Pauli* | Theodor Petra, ed., *PL* 152:9–567. | 283,797 |
| *Bruno Carthusianorum* | *Expositio in Psalmos* | Ibid., *PL* 152:637–1420. | 414,731 |
| | | | 698,528 |
| Gerhoh of Reichersberg (1093–1169) | *Commentarius aureus in Psalmos et cantica ferialia* | Bernhard Pez, ed., *PL* 193:619–1814. | 594,011 |
| *Gerhohus Reicherspergensis* | *De aedificio dei* | Ibid., *PL* 194:1187–1334. | 54,343 |
| | *Epistolae Gerhohi* | Ibid., *PL* 193:489–618. | 64,584 |
| | | | 712,938 |
| Gilbert of Poitiers (1075–1154) | *Expositio in Boethii librum De bonorum hedbomade* | Nikolaus K. Häring, ed. | 13,013 |
| *Gislebertus Porretanus* | *Liber de sex principiis* | Hermolaus Barbarus, ed., *PL* 188:1257–70. | 4,593 |
| | | | 17,606 |
| Guibert de Nogent (c. 1053–1125) | *Carmen Erumnarum et dolorum* | Robert B.C. Huygens, ed., *CC CM* 127A, 371–3. | 570 |
| *Guibertus de Novigento* | *Contra iudaizantem et Iudaeos* | Ibid., 316–73. | 14,697 |
| | *De bucella Iudae data* | Ibid., 65–77. | 3,257 |
| | *De sanctis et eorum pigneribus* | Ibid., 79–175. | 26,138 |
| | *De vita sua sive Monodiae* | ed. by Edmond-René Labande, *Les classiques de l'histoire du Moyen Âge* 34. | 47,171 |
| | *Historia quae inscribitur Dei gesta per Francos* | Robert B.C. Huygens, ed., *CC CM* 127A, 77–352. | 55,363 |
| | *In Zachariam* | Ibid., 367–70. | 998 |
| | *Quo ordine sermo fieri debeat* | Ibid., 47–63. | 4,401 |
| | | | 152,585 |
| Hildebert of Lavardin (c. 1055–1133) | *Epistolae de Paschali papa* | Ernst Sackur, ed., *MGH*, *Ldl* 2, 668–73. | 1,954 |
| *Hildebertus Lavardinensis* | *Vita Mariae Aegyptiacae* | Norbert Klaus Larsen, ed., *CC CM* 209. | 6,043 |
| | | | 7,997 |
| Honorius of Regensburg (c. 1080–c. 1154) | *De apostatis* | Julius Dieterich, ed., . *MGH, Ldl 3,* 57–63. | 2,079 |
| *Honorius Augustodunensis* | *De offendiculo* | Ibid., 38–57. | 7,485 |
| | *Elucidarium* | John Allen Giles, ed., *PL* 172:1109–76. | 31,785 |
| | *Expositio in Cantica canticorum* | Philippe Despont, ed., *PL* 172:347–496. | 75,588 |
| | *Gemma animae* | Melchior Lotter, ed., repr. in *PL* 172:541–738. | 86,102 |

Table A.I.1: Benchmark corpus of eleventh- and twelfth-century Latin texts.

| Author | Texts | Edition | *n* words |
|---|---|---|---|
| | *Imago mundi* | Philippe Despont, ed., *PL* 172:115–88. | 27,069 |
| | *Imago mundi: Continuatio Weingartensis* | Ludewico Weiland, ed., *MGH SS* 43 (rer. Germ.), 61–3. | 744 |
| | *Imago mundi (cum continuationibus VII) (Excerpta)* | Roger Wilmans, ed., *MGH SS* 10 (*Scriptores*), 132–34. | 740 |
| | *Speculum ecclesiae* | Jacques-Paul Migne, ed., *PL* 172:807–1109. | 136,812 |
| | *Summa gloria* | Julius Dieterich, ed., *MGH, Ldl 3,* 63-80. | 5,377 |
| | *Summa totius* | Wilmans, ed., *MGH SS* 10 (*Scriptores*), 128-31. | 2,926 |
| | | | 376,697 |
| Hugh of Saint-Victor (c. 1096–1141)<br>*Hugo de Sancto Victore* | *De Sacramentis Christianae fidei* | The canons regular of Saint-Victor, ed., (Rouen, 1648), *PL* 176:173–618. | 185,015 |
| | *De septem donis spiritus sancti* | Roger Baron, ed., *SC* 155, 120–32. | 1,318 |
| | *De substantia dilectionis* | Ibid., 82–92. | 1,219 |
| | *De tribus rerum subsistentiis* | Charles Henry Buttimer, ed., *Fontes Christiani* 27, 134–5. | 313 |
| | *De vanitate rerum mundanarum* | Cédric Giraud, ed., *CC CM* 269, 137–200. | 14,487 |
| | *De verbo dei* | Roger Baron, ed., *SC* 155, 60–80. | 2,405 |
| | *Dialogus de creatione mundi* | Cédric Giraud, ed., *CC CM* 269, 299–348. | 10,000 |
| | *In hierarchicam coelestem* | The canons regular of Saint-Victor, ed. (Rouen, 1648), *PL* 175:923–1154. | 89,005 |
| | *Sententiae de divinitate* | Ambrogio M. Piazzoni, ed., *Studi Medievali* 23, 912–55. | 16,805 |
| | | | 317,581 |
| Ivo of Chartres (1040–1115)<br>*Ivo Carnotensis* | *Decretum* | Jean Fronteau, ed., *PL* 161:9–1021. | 395,190 |
| | *Epistolae ad litem investiturarum spectantes* | Ernst Sackur, ed., *Ldl* 2. *MGH*, 642–57. | 4,293 |
| | *Panormia* | Melchior Vosmedian, ed., rev. by Jacques Paul-Migne. *PL* 161:1037–1344. | 111,884 |
| | | | 511,367 |
| John of Salisbury (c. 1125–1180)<br>*Ioannes Saresberiensis* | *Epistolae* | John Allen Giles, ed., *PL* 199:1–379. | 169,324 |
| | *Metalogicon* | Ibid., *PL* 199:823–944. | 59,273 |
| | *Polycraticus* | Ibid., *PL* 199:379–822. | 214,061 |
| | *Vita Sancti Anselmi Cantuariensis* | Ibid., *PL* 199:1009–1040. | 13,809 |
| | | | 456,467 |
| Peter Damian (c. 1007–c. 1073)<br>*Petrus Damianus* | *De divina omnipotentia* | Constantino Cajetan, ed., *PL* 145:595–622. | 13,118 |
| | *Liber Gomorrhianus* | Ibid., *PL* 145:159–90. | 14,157 |
| | *Vita sancti Romualdi* | Ibid., *PL* 144:953–1008. | 18,764 |
| | | | 46,039 |
| Peter Lombard (c. 1096–1160)<br>*Petrus Lombardus* | *Collectanea in epistolas Pauli* | Josse Bade, ed., *PL* 191:1297–192:520. | 264,743 |
| | *Commentaria in Psalmos* | Richard of Le ed., Mans, *PL* 191:55–1296. | 661,137 |
| | *Sententiae* | Jean Aleaume, ed., *PL* 192:519–964. | 293,155 |
| | | | 1,219,035 |
| Peter of Celle (c. 1115–1183)<br>*Petrus Cellensis* | *De conscientia* | Ambroise Janvier, ed., *PL* 202:1083–98. | 5,817 |
| | *Epistolae* | Ibid., *PL* 202:405–636. | 100,501 |
| | *Liber de panibus* | Ibid., *PL* 202:929–1046. | 58,071 |
| | *Mystica et moralis expositio Mosaici tabernaculi* | Ibid., *PL* 202:1047–84. | 17,826 |
| | *Sermones* | Ibid., *PL* 202:637–928. | 141,491 |
| | | | 323,716 |
| Peter the Venerable (1092–1156)<br>*Petrus Venerabilis* | *Adversus Iudaeorum inveteratam duritiem* | Yvonne Friedman, ed., *CC CM* 58. | 60,230 |
| | *Adversus sectam Saracenorum* | Edmond Martène, ed., *PL* 189:659–720. | 29,640 |

Table A.I.1: Benchmark corpus of eleventh- and twelfth-century Latin texts.

| Author | Texts | Edition | *n* words |
|---|---|---|---|
| | *Carmina* | Ibid., *PL* 189:1005–24. | 6,834 |
| | *Contra Petrobrusianos haereticos* | James Fearns, ed., *CC CM* 10. | 52,704 |
| | *De miraculis libri duo* | Dénise Bouthillier, ed., *CC CM* 83. | 38,282 |
| | *Epistulae* | Giles Constable, ed., *Harvard Historical Studies* 78. | 142,262 |
| | | | 329,952 |
| Rupert of Deutz (c. 1075–1130) | *Anulus sive dialogus inter Christianum et Iudaeum* | Rabanus Haacke, ed., Arduini, "Ruperto di Deutz," | 20,939 |
| *Rupertus Tuitiensis* | *Carmina de calamitatibus ecclesiae Leodiensis* | Justus Henning Böhmer, ed., *MGH*, *Ldl* 3, 624–41. | 3,025 |
| | *Commentaria in Canticum Canticorum* | Rabanus Haacke, ed., *CC CM* 26. | 53,510 |
| | *Commentaria in duodecim prophetas minores* | Michael Pleunich, ed., *PL* 168:9–863. | 354,235 |
| | *Commentaria in evangelium sancti Iohannis* | Rabanus Haacke, ed., *CC CM* 9. | 266,430 |
| | *Commentarium in Apocalypsim Iohannis apostoli* | Michael Pleunich, ed., *PL* 169:825–1214. | 163,647 |
| | *De gloria et honore* | Rabanus Haacke, ed., *CC CM* 29, 3–421. | 146,854 |
| | *De glorificatione Trinitatis et processione Spiritus sancti* | Michael Pleunich, ed., *PL* 169:9–202. | 79,366 |
| | *De incendio Tuitiensi* | Herbert Grundmann, ed., *Deutsches Archiv* 22, 441–71. | 9,683 |
| | *De meditatione mortis* | Michael Pleunich, ed., *PL* 170:357–90. | 13,681 |
| | *De sancta trinitate et operibus eius* | Rabanus Haacke, ed., *CC CM* 21–4. | 681,271 |
| | *De victoria verbi Dei* | Ibid., *MGH QQ zur Geistesgesch.* 5, 1–246. | 117,392 |
| | [?] *Epistula ad F.* | J. Lajos Csóka, ed., *Studien und Mitteilungen zur Geschichte des Benediktiner-Ordens* 84, 383–5. | 603 |
| | *Hymnus primus de sancto spiritu* | Rabanus Haacke, ed., *CC CM* 29, 380–1. | 101 |
| | *Hymnus secundus sive Oratio ad sanctum spiritum* | Ibid., 422–4. | 347 |
| | *Liber de divinis officiis* | Ibid., *CC CM* 7. | 128,157 |
| | *Liber de laesione virginitatis* | Michael Pleunich, ed., *PL* 170:545–60. | 5,658 |
| | [?] *Officium de festo Sancti Augustini* | In *Thomae Aquinatis opera omnia* 24, 237–40. | 1,902 |
| | *Passio Eliphii Tullensis* | Michael Pleunich, ed., *PL* 170:427–36. | 3,659 |
| | *Sermo de Pantaleone* | Maurice Coens, ed., *Analecta Bollandiana* 55, 254–67. | 3,902 |
| | [?] *Vita Heriberti Coloniensis* | Peter Dinter, ed., *Veröffentlichungen des Historischen Vereins für den Niederrhein* 13, 30–88. | 14,942 |
| | | | 2,069,304 |
| Walter of Châtillon (c. 1135–c. 1179) | *Alexandreis* | Athanasius Gugger, ed., *PL* 209:459–572. | 44,313 |
| *Gualterus de Castellione* | [?] *De SS. Trinitate tractatus* | Bernhard Pez, ed., *PL* 209:573–90. | 7,230 |
| | *Tractatus contra Iudaeos* | Casimir Oudin, ed., *PL* 209:423–458. | 17,875 |
| | | | 64,418 |
| William of Conches (c. 1090–1154) | *Dragmaticon Philosophiae* | Italo Ronca and Leonard F. Badia, ed., *CC CM* 152. | 42,191 |
| *Guillelmus de Conchis* | *Glosae super Boetium* | L. Nauta, ed., *CC CM* 158. | 61,697 |
| | | | 103,888 |
| William of Saint-Thièrry (c. 1080–1148) | *Aenigma fidei* | Bertrand Tissier, ed., *PL* 180:397–440. | 22,841 |
| *Guillelmus de Sancto Theodorico* | *Brevis commentatio (in Cantici Canticorum priora dua capita)* | Jean Mabillon, ed., *PL* 184:407–36. | 14,822 |
| | *De contemplando Deo* | Jean Mabillon, ed., *PL* 184:365–79. | 7,721 |
| | *Epistola de erroribus Guillelmi de Conchis* | Bertrand Tissier, ed., *PL* 180:333–340. | 3,917 |
| | *De natura corporis et animae* | Ibid., *PL* 180: 695–726. | 15,041 |
| | *De natura et dignitate amoris* | Jean Mabillon, ed., *PL* 184:379–408. | 13,753 |
| | *De sacramento altaris* | Bertrand Tissier, ed., *PL* 180:341–66. | 11,643 |
| | *Disputatio adversus Abaelardum* | Ibid., *PL* 180:249–82. | 16,925 |

Table A.I.1: Benchmark corpus of eleventh- and twelfth-century Latin texts.

| Author | Texts | Edition | *n* words |
|---|---|---|---|
| | [?] *Disputatio altera adversus Abaelardum* | Ibid., *PL* 180:283–328. | 24,105 |
| | *Expositio altera super Cantica canticorum* | Ibid., *PL* 180:473–545. | 37,605 |
| | *Expositio in Epistolam ad Romanos* | Ibid., *PL* 180:547–694. | 75,224 |
| | *Orationes meditativae* | Ibid., *PL* 180:205–48. | 23,160 |
| | *Speculum fidei* | Ibid., *PL* 180:365–98. | 16,626 |
| | | | 238,100 |
| **Total** | | | 6,985,668 |

# A.2 Addenda to Chapter 2

| *sample size* | *Yule's* K |
|---|---|
| 100 words | 48.0 |
| 200 words | 97.5 |
| 300 words | 44.0 |
| 400 words | 38.88 |
| 500 words | 31.76 |
| 600 words | 45.89 |
| 700 words | 40.33 |
| 800 words | 41.19 |
| 1,000 words | 37.7 |
| 2,000 words | 41.24 |
| 3,000 words | 43.06 |
| 4,000 words | 42.56 |
| 5,000 words | 50.05 |
| 6,000 words | 46.51 |
| 7,000 words | 37.31 |
| 8,000 words | 41.82 |
| 10,000 words | 39.06 |
| 20,000 words | 45.41 |
| 30,000 words | 46.86 |
| 40,000 words | 46.25 |
| 50,000 words | 41.03 |
| 60,000 words | 45.13 |
| 70,000 words | 46.33 |
| 80,000 words | 45.35 |
| 100,000 words | 45.95 |
| 200,000 words | 45.8 |
| 300,000 words | 45.53 |

(a) Data for fig. 2.2a: Yule's $K$. Note the constancy of the results despite the differing sample length.

| *feature* | $\frac{f}{N}$ | *feature* | $\frac{f}{N}$ |
|---|---|---|---|
| et | 4.3% | nec | 0.28% |
| in | 2.84% | me | 0.27% |
| est | 2.29% | ab | 0.26% |
| non | 1.58% | etiam | 0.26% |
| quod | 1.05% | sic | 0.25% |
| ut | 0.98% | pro | 0.24% |
| ad | 0.95% | sit | 0.23% |
| qui | 0.86% | ergo | 0.22% |
| de | 0.85% | ita | 0.22% |
| que | 0.79% | quo | 0.21% |
| cum | 0.72% | quid | 0.19% |
| sed | 0.7% | hec | 0.19% |
| quia | 0.67% | aut | 0.19% |
| per | 0.59% | te | 0.18% |
| hoc | 0.54% | eum | 0.18% |
| id | 0.53% | uero | 0.18% |
| a | 0.51% | deo | 0.18% |
| si | 0.5% | tamen | 0.18% |
| sunt | 0.49% | super | 0.18% |
| uel | 0.48% | quasi | 0.18% |
| enim | 0.43% | secundum | 0.17% |
| eius | 0.38% | scilicet | 0.17% |
| autem | 0.37% | eo | 0.17% |
| dei | 0.36% | ipse | 0.17% |
| quam | 0.36% | iam | 0.16% |
| ex | 0.35% | dominus | 0.16% |
| esse | 0.33% | unde | 0.15% |
| se | 0.29% | nisi | 0.15% |
| sicut | 0.29% | ne | 0.15% |
| deus | 0.28% | eorum | 0.15% |

(b) Data for fig. 2.2b: Zipf's law. $\frac{f}{N}$ denotes the relative frequency.

Table A.2.1: Data for fig. 2.2.

# A.3 Addenda to Chapter 3

## A.3.1 Coding and Programming

Those who are completely unfamiliar with coding and programming might be grateful to learn about it in somewhat more detail. The current section means to give only a concise introduction.

### Commands and Programming Languages

A computer runs on an operating system (OS), a kind of dominant software that steers and manages the computer hardware (the physical device). In intuitive terms the OS forms the brain of a computer, whereas the hardware forms the body. There are different operating systems and different devices imaginable, and some of them are transferable to other devices (as is the popularly used Linux, for example). In essence, everything a person ever does on his or her computer is launch commands directed towards the OS. The execution of a command entails a number of calculations made, which results in the completion of a desired query. Coding and programming are no condition in order to launch commands at a computer. Also those who are unfamiliar with explicit coding and programming will launch commands to the OS, be it that the commands are given through a user-friendly graphical user interface (GUI). This means that, instead of explicating commands through coding and programming, most users will simply be clicking and scrolling through applications, processors and browsers on a computer, which do most of the command-giving towards the OS for them behind the screens. The flexibility of the GUI can become, however, rather limited when one wishes to execute a more complex series of calculations with a personal set of data, and this is where coding becomes important.

Coding and programming allow for a more direct interaction with the computer. The idea is that one writes up a script of actions (a set of instructions) that have to be undertaken by the computer, in a specific programming language with a specific syntax (i.e. rules that define meaningful combinations of inserted symbols). This can be done both interactively in the command line (also commonly referred to as the Shell), which is usually done for launching simple commands and blocks of code, or else in script mode, which provides for a long series of specific commands which the computer runs through line by line. There are hundreds of programming languages (e.g. Python, Java, Javascript, C, C++). Although which of these is the superior language might well be one of programmers' favorite topics of debate, there is in fact no consensus as to which language is preferable over another. This would be a matter of convention or personal preference, or slightly dependent on the task at hand —although here, again, one can easily disagree over which programming language fits which task best. There are some languages that have sought up the boundaries between programming languages and natural languages by —either jokingly or for

```
#!/usr/bin/env

a = 2
b = 3
if a == 2:
^^Iprint('medieval literature')
```

Listing 1: Intuition of programming logic.

the sake of aesthetics— developed so-called 'esoteric' programming languages.[1] The results of such experimenting can go quite far, to the extent that some coding languages have a syntax in which one is required to code as if one was writing a Shakespeare play, indeed, in acts. Interestingly, also programming languages have been shown to exhibit a personal style, much like natural languages,[2] which makes one marvel about a possible link between a person's speech and a person's logical reasoning. Despite the exceptional esoteric languages, which tend to be highly meta, the syntax of all programming languages is practically founded upon highly similar logical principles, so that in essence all languages can be applied for most programming tasks. See, for instance, the simple block of code below, written in `Python`, where the computer is asked to execute a command (i.e. return the string `'medieval literature'`) if the condition is met (`if`-condition) that the variable `a` equals 2:

The script snippet above is a straightforward example of programming logic. Firstly, one declares variables and stores information in them, which corresponds to our data (or input). Consequently, one can analyze and handle the data by steering the computer through commands, such as here: if a requirement is met (i.e. if it is either `True` or `False`), then the computer will return (print to the screen) a string (a sequence of characters) which says `'medieval literature'`, our output.

The example above solely means to provide an intuition of the dynamics and implications that come with processing data through a computer. As one will be able to observe, it constitutes a highly simple series of commands, almost a basic calculator task, which is easy for the computer to solve and takes but milliseconds to run. When programmers write software, however, numerous of such conditions and decision functions such as the one above will be chained up. They will write up to hundreds and thousands of lines of code in order to construct sophisticated algorithms, chains of step-by-step functions which output decisions, which can in its own turn be interpreted again, etc. Usually one will come across the term 'model,' because what a programmer does is in fact model or tailor stacks of lines of codes so that they will

[1] Michael Mateas and Nick Montfort, "A Box, Darkly: Obfuscation, Weird Languages, and Code Aesthetics," in *Proceedings of the 6th Annual Digital Arts and Culture Conference* (Copenhagen, Denmark, December 2005).

[2] Aylin Caliskan-Islam et al., "De-anonymizing Programmers via Code Stylometry," in *Proceedings of the 24th USENIX Security Symposium* (Washington, D.C., August 2015), 255–70.

be able to solve a specific problem. In that light, this is also the place to emphasize that coding is often perceived as a collective endeavour, where lines and bits of code are freely shared and encouraged to be openly accessible so that a community of programmers will be able to use them. In that sense, a programmer should not be writing everything from scratch. (S)he will be able to draw from so-called language-specific distributions, or else draw on the numerous modules (single scripts), packages (collection of modules) or libraries (collection of packages) that can be found online, for example such as in specialized repositories as Github. Importing such pre-written bits of software is time-economic, and often allows for programs to run much faster. Another advantage is that these are built by expert programmers, and are well-documented. To give but one example, most of the plots and figures in this thesis are generated by a package called Matplotlib.[3]

### Natural Language Processing

For the larger part one will find that problems in computer science will be of a mathematical nature, as the environment is evidently built upon mathematical principles. When dealing with texts, however, one descends into a subfield of computer science called natural language processing (NLP). NLP comes with the implication that texts are computationally processed, i.e. they are dropped into an environment that operates onto mathematical principles. Whereas our variables `a` and `b` in the code listing above contained mathematical information, the NLP scholar processes variables containing 'strings' (i.e. characters, words, sentences, or entire texts). The idea is consequently to process, transform and mine electronic text through numerical methods. The possibilities of NLP are not confined to stylistic analysis only. Other well-known examples are semantic analysis ("determining the meaning of character sequences or word sequences"), sentiment analysis ("determining the sentiments behind a character sequence"), information retrieval ("the process of retrieving information [...] corresponding to a query that has been made by the user"), discourse analysis ("the process of performing text or language analysis, which involves text interpretation and knowing the social interactions"), and many other applications.[4] Aside from —naturally— computational stylistics itself, this thesis will be little concerned with the other aforementioned types of NLP. Exceptions are the basic operations, the so-called preprocessing tasks of NLP that lead up to the possibility of performing stylistic analysis. Such basic operations include for instance tokenization, normalization, calculating word frequencies, etc. These terms will become more clear as we proceed throughout this chapter, but their simple power can be illustrated here already:

[3] John D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering* 9, no. 90 (2007): 90–5.

[4] Deepti Chopra, Nisheeth Joshi, and Iti Mathur, *Mastering Natural Language Processing with Python* (Birmingham, UK: Packt Publishing, 2016), definitions respectively found on 107, 133, 165 and 183.

```
1  In [1]: from collections import Counter
2  In [2]: string = 'Itaque scribe ista non secundum cor tuum, sed secundum testimonium
   ↪  meum, qui sine initio et fine uita sum, nec per te inuenta nec per alium hominem
   ↪  premeditata, sed per me ante principium mundi preordinata; quoniam ut ante
   ↪  creatum hominem ipsum presciui, sic etiam illa que ei necessaria sunt preuidi.'
3  In [3]: string.split()
4  Out[3]: ['Itaque,' 'scribe,' 'ista,' 'non,' 'secundum,' 'cor,'  'tuum,,' 'sed,'
   ↪  'secundum,' 'testimonium,' 'meum,,' 'qui,'  'sine,' 'initio,' 'et,' 'fine,'
   ↪  'uita,' 'sum,,'  'nec,' 'per,' 'te,' 'inuenta,' 'nec,' 'per,'  'alium,'
   ↪  'hominem,' 'premeditata,,' 'sed,' 'per,' 'me,'  'ante,' 'principium,' 'mundi,'
   ↪  'preordinata;,' 'quoniam,' 'ut,'  'ante,' 'creatum,' 'hominem,' 'ipsum,'
   ↪  'presciui,,' 'sic,'  'etiam,' 'illa,' 'que,' 'ei,' 'necessaria,' 'sunt,'
   ↪  'preuidi.']
5  In [4]: Counter(string).most_common(2)
6  Out[4]: [('per,' 3), ('secundum,' 2)]
```

Listing 2: Intuition of basic NLP task scripted in Python, where a string of text is segmented (split) in separate tokens contained in a list, counted and ordered by frequency. The two most common terms are outputted through the command in line 5.

In the listing above, first a container type called `Counter` is imported from a module called `collections`, which is part of the standard Python Standard Library (line 1). Consequently, a sentence by Hildegard of Bingen is stored as a variable which we have called `string` (line 2). This string is subsequently split into separate words. After this, a container type `Counter` is summoned, which can be tallied for the two most common words. Even those unfamiliar with Python can probably get a good grasp of what is going on in this early stage. The analyst loads a text into a project (line 2), consequently the analyst decides upon the level of the text which (s)he would wish to focus on (line 4: here the string was segmented on the word level, but (s)he could have also opted for characters, so-called *n*-grams or sentences),[5] and consequently a very easy algorithmic operation is performed which outputs a tallied result: `'per'` and `'secundum'` are the two most commonly occurring words within this sentence of Hildegard of Bingen, by a respective frequency of 3 vs. 2 times. There we have it: a distant reading of a single sentence by Hildegard of Bingen.

Throughout this thesis we will seldom return to the raw code underlying the computational experiments (which have all been written in Python). Only the very vital elements within the experimental set-up will be reported. I thereby hope to strike a balance between transparency and economy of space. In any case, the source code of all experiments are consultable on Github, if one is interested to gain a fuller understanding of the steps involved.

[5] This process is commonly called 'tokenization,' we return to this on p. 313.

### A.3.2 Preprocessing: Formatting Latin

In code listing 2 above, one might have already noticed that a number of potential problems arise when a text is analyzed in its raw form. Punctuation, for example, was retained. Upper cases are still present. Texts have to be explicitly normalized for a computer in order to avoid superfluous duplicates, which can considerably problematize the subsequent steps of an experiment. In this section, we therefore address what is often perceived as the most important step in any data analysis project: preprocessing, i.e. preparing and cleaning the data prior the experiment. The state of the underlying text is vital. As a popular data analysis quote goes: "garbage in, garbage out" (acronym: GIGO). The quality of the underlying corpus is essential in any experiment. If the data is flawed, so will be the results. Therefore, any nonchalance in the prior steps of the experiment will be catastrophic, in the sense that the results might be false or heavily biased.[6]

Commonly one might find that data is 'dirty,' 'noisy' or 'damaged.' The degree of such damage, then again, heavily depends on the origin of the data, and judging whether or not data is indeed unacceptably dirty or not is dependent on the nature of the experiment. For example, the analyst will have to ask him- or herself whether elements of mark-up —if present in the text— such as paragraph numbers, brackets, chapter titles, footnotes, etc. are significant to the task at hand. In some situations they might well be. If not, one would be well advised to remove them. More drastic examples of situations in which data can be dirty are, for example, if the data was a result of large-scale OCRing (optical character recognition) or HTRing (handwritten text recognition): the automatic recognition of respectively scanned or typeset print on the one hand and handwritten text on the other. Such automated recognition tasks can come with a considerable margin of error. In essence, preprocessing boils down to a combination of demarcating and redefining the data on the desired level of analysis, i.e. splitting it up in tokens through *tokenization*, and securing the quality of the token by applying *normalization*. The order in which the aforementioned processes occur can differ accordingly.

#### Tokenization

Tokenization is essentially just another term for an intuitive given that should by now be somewhat familiar. In code listing 2 above (p. 312), for example, we have already witnessed an example (line 4) in which a string of text was segmented into tokens. The idea is that one defines a separator character (e.g. a space character) or a separator rule by which an input string of text is transformed into a sequence of meaningful sections. There are numerous versions upon which one can slice up language into sections:

[6] Maciej Eder, "Mind Your Corpus: Systematic Errors in Authorship Attribution," *Literary and Linguistic Computing* 28, no. 4 (2013): 603–14.

```
1  # original string
2  string = 'Cum enim Deus primum angelum, qui Lucifer dictus est, cum omnibus
   ↪  ornamentis creaturarum, que omnibus creaturis dederat, ita ornasset, ut etiam
   ↪  totum agmen eius inde splendorem haberet'
3  # split on the comma
4  ['cum enim Deus primum angelum,' 'qui Lucifer dictus est,' 'cum omnibus ornamentis
   ↪  creaturarum,' 'que omnibus creaturis dederat,' 'ita ornasset,' 'ut etiam totum
   ↪  agmen eius inde splendorem haberet']
5  # split on character bigrams
6  char_bigrams = [('c,' 'u'), ('u,' 'm'), ('m,' ' '), (' ,' 'e'), ('e,' 'n'), ('n,'
   ↪  'i'), ('i,' 'm'), ('m,' ' '), (' ,' 'd'), ('d,' 'e'), ('e,' 'u'), ('u,' 's'),
   ↪  ('s,' ' '), (' ,' 'p'), ('p,' 'r'), ('r,' 'i'), ('i,' 'm'), ('m,' 'u'), ('u,'
   ↪  'm'), ('m,' ' '), (' ,' 'a'), ('a,' 'n'), ('n,' 'g'), ('g,' 'e'), ('e,' 'l'),
   ↪  ...]
7  # split on word bigrams
8  word_bigrams = [('cum,' 'enim'), ('enim,' 'deus'), ('deus,' 'primum'), ('primum,'
   ↪  'angelum,'), ('angelum,,' 'qui'), ('qui,' 'lucifer'), ('lucifer,' 'dictus'),
   ↪  ('dictus,' 'est,'), ('est,,' 'cum'), ('cum,' 'omnibus'), ('omnibus,'
   ↪  'ornamentis'), ('ornamentis,' 'creaturarum,'), ('creaturarum,,' 'que'), ('que,'
   ↪  'omnibus'), ...]
```

Listing 3: Intuition of tokenization.

As demonstrated in the listing above, tokenization can take place at several levels: the character, word, phrase and sentence level. Consequently, once the basic token is defined, one can further customize sequences in $n$ ranges ($n$ denoting a variable integer[7]), resulting into so-called $n$-grams, sequences of $n$ tokens from a given sample of text or speech. If $n$ equals 2 or 3, one will respectively encounter the terms bigram and trigram. Tokenization is a process by which textual data becomes manageable in a new form, a form upon which the analyst decides. For stylistic analysis, specifically, this means that the analyst has —from the very outset— quite a few options to choose from. Note that the code listing serves merely as an illustration. I do not necessarily subscribe to the idea that tokenizing on the comma (as illustrated in line 4), is a good idea in the case of medieval Latin. The discussion in the preceding chapter on sentence length should have already indicated the problems attached to considering phrases / sentences as units possessing stylistic information. We will qualitatively review such decisions at a later stage in this chapter.

### Normalization

Token normalization is the process of transformation of the original data to a standard, indeed, a norm to which all individual texts within the dataset comply. Normalization

[7] An integer is a whole number that is not a fraction, as opposed to a float number.

is, therefore, also an establishment of alignment and compatibility between data of varying origins. As can already be observed in listing 3, on the one hand normalization can imply, for example, basic preprocessing tasks such as the deletion of superfluous punctuation, or the replacement of upper case letters by their lower cases. However, dependent on the study subject, more extensive, data-specific normalization operations can surface. For instance, when it comes to comparing two medieval Latin texts, one can bump into a rich variety of orthographic and editorial conventions,[8] which are best aligned in order to avoid duplicates. Some examples are:

- <j>'s vs. <i>'s
- <ae>'s vs. <e>'s, or <oe>'s vs. <e>'s
- <v>'s vs. <u>'s
- lenition: *racio* vs. *ratio* or *multociens* vs. *multotiens*
- strengthened aspiration or fortition, e.g. *michi* vs. *mihi* (other examples: *nichil* vs. *nihil*)
- progressive or regressive assimilation / dissimilation: *exsistere* vs. *existere*, *obf–* vs. *off–* (e.g. *obfuscare*), *abji-* vs. *abi-* (e.g. *abjiciendus est*), *adm–* vs. *amm–* (e.g. *ammonitio*, *ammiratio*), etc.
- *quandiu* vs. *quamdiu* (other examples are: *tanquam* vs. *tamquam*, *nunquam* vs. *numquam*, *-cunque* vs. *-cumque*, ...), *tandiu* vs. *tamdiu* (and *quandiu* vs. *quamdiu*)
- *quatenus* vs. *quatinus*
- *imo* vs. *immo*
- *quoties* vs. *quotiens* (other examples: *totiens* vs. *toties*)
- *velut* vs. *velud*

Mostly, normalization can be executed on the basis of rule-based, fully automated operations. A possible means of doing so is by using regular expressions (often denoted as `regex` or `regexp`). Such regular expressions are powerful and flexible methods in which sequences of characters define word search patterns. They are often applied for 'find' or 'find and replace' operations, a more complex version of the 'find and replace' methods encountered in text processors. One could also think of `regex` as a specific type of language, which can be used to precisely trace and edit specific patterns within words.[9] Although `regex` is powerful, it can easily reach its limits, and it is not very time-economic: there are more sophisticated algorithmic procedures of normalization, which will be discussed below.

[8] Arthur G. Rigg, "Orthography and Pronunciation," in *Medieval Latin : An Introduction and Bibliographical Guide*, ed. Frank A.C. Mantello and Arthur G. Rigg (Washington, D.C.: The Catholic University of America Press, 1996), 79–82.

[9] Steven Bird, Ewan Klein, and Edward Loper, *Natural Language Processing with Python* (Sebastopol: O'Reilly, 2009), 97ff.

### Disambiguation: Stemming, Lemmatization, PoS-tagging and Parsing

Normalization ranges from simple 'find and replace'-tasks to more sophisticated tasks. For instance, it might well be possible (dependent on the task at hand) that an analyst might like to strip off affixes or enclitics (just think of the enclitic conjunction *-que* in Latin), or that an analyst would like to disambiguate between homographs such as, again, the enclitic *-que* and the relative pronoun *quae* (which can occasionally be written as *que*). For Latin, such tasks can be particularly challenging. Many Latin suffixes cannot be automatically linked to an unambiguous morphological category. Words ending in *–ter*, for example, correspond to no less than six different parts of speech: nouns (*fra-ter*), adjectives (*dex-ter*), pronouns (*al-ter*), adverbs (*gravi-ter*), numeral adverbs (*qua-ter*) and prepositions (*in-ter*). Additionally, like many other languages, Latin is teeming with homographs which require context to be disambiguated. A token such as *legi* can both be lemmatized under the verb *lego* as under the noun *lex*. Similarly ambiguous tokens include common forms such as *quae*, *satis* or *venis*. For lemmatization specifically, another problem is verb forms which show no resemblance to their lemma. The fact that *tuli* is an active 1st person singular perfect of *fero* is not obvious, and the same problem applies to *fero*'s perfect participle *latus*, which could in its own turn be confused with the homonymous common noun *latus* (transl. 'side').

Simple find and replace tasks, then, are insufficient to deal with the complexity of the language. A tagger has to learn, for example, the morphological connection between *tuli*, *latus* and *fero* by moving beyond superficial outward appearances (prefixes, word stems or suffixes), and by properly modelling the immediate context surrounding these words.[10] For normalization to make such advanced decisions, one needs more sophisticated NLP software such as stemmers, lemmatizers, part-of-speech (PoS) taggers and parsers:

- **Stemmer**: a tool that has the purpose of automatically distinguishing word stem, base or root form from morphological affix:
  `serui serv|i`
- **Lemmatizer**: automatically refers a token to its headword (or lemma) as it can be consulted in a dictionary:
  `serui servus`
- **Part-of-speech tagger**: a PoS-tagger comes in a *simplex* and *complex* variation, and attempts to automatically determine the part of speech of a token, and if possible, specific morphological information of that token. The *simplex* variant is restricted to the part-of-speech as such:

[10] Much of the foregoing is literally taken from a publication I have written in co-authorship with Mike Kestemont, which goes into somewhat more detail, see Mike Kestemont and Jeroen De Gussem, "Integrated Sequence Tagging for Medieval Latin Using Deep Representation Learning," ed. Marco Büchler and Laurence Mellerin, *Journal of Data Mining and Digital Humanities*, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages, 2017, 1–2, `https://arxiv.org/abs/1603.01597`.

```
serui NN
```

The *complex* variant, on the other hand, takes into account person, number, gender, case, or —in case of verbs— information such as voice, mood and tense:

```
exhortationis NN case=GENITIVE gender=FEMININE number=SINGULAR
scripserim V mood=SUBJUNCTIVE number=SINGULAR person=PERSON_1
tense=PERFECT voice=ACTIVE
```

- **Parser**: Tags the syntactic information of tokens, and works out the grammatical structure of sentences. Parsers will often output structure trees or dependencies to denote grammatical relations. For instance, the parser will automatically work out the grammatical function of a token, i.e. whether or not a certain noun is the subject or object of a phrase.

For Latin, there are quite a few projects concerned with enabling some or all of the above, by establishing annotated corpora and building upon these resources by the further supply of software packages (open-source or accessible through payment). Again, the manual annotations by pioneer Roberto Busa SJ and his *Index Thomisticus* provided the groundworks for the technical development of additional tools.[11] Further work on automatic normalization, lemmatization and morpho-syntactic annotation of Latin has been further explored by the LASLA center in Liège,[12] the *Perseus* Project at Boston's Tufts University,[13] LatinISE in Cambridge, UK,[14] PROEIL in Oslo (*Pragmatic Resources of Old Indo-European Languages*),[15] the European CHLT project stationed for most part in Milan and Pisa (*Cultural Heritage Language Technologies*)[16] and the CompHistSem team in Frankfurt (*Computational Historical Semantics*).[17] Here I lack the space to discuss these projects' scope in full. I refer any reader interested in them to a publication where I have been more thorough in describ-

[11] Marco Carlo Passarotti, "One Hundred Years Ago. In Memory of Father Roberto Busa SJ.," in *Proceedings of The Third Workshop on Annotation of Corpora for Research in the Humanities (ACRH-3)*, ed. Francesco Mambrini, Marco Passarotti, and Caroline Sporleder (Sofia, Bulgaria, December 2013), 15–24.

[12] Their list of publications is extensive and too much to cite here. See http://web.philo.ulg.ac.be/lasla/publications/

[13] The project began in 1987, and the first release of Latin texts on Perseus came in 1997, see Crane and Mylonas, "The Perseus Project."

[14] LatinISE itself is a text corpus annotated with NLP tools. It was assembled by McGillivray by drawing from the *LacusCurtius*, *Intratext* and *Musisque Deoque* websites, in collaboration with the late Adam Kilgarriff († 2015) of the corpus query tool Sketch Engine, see Barbara McGillivray and Adam Kilgarriff, "Tools for Historical Corpus Research, and a Corpus of Latin," in *New Methods in Historical Corpora*, ed. Paul Durell et al., vol. 3, Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache (Tübingen: Narr, 2013).

[15] Dag Trygve Truslew Haug and Marius Larsen Jøhndal, "Creating a Parallel Treebank of the Old Indo-European Bible Translations," in *Proceedings of the Second LREC Workshop on Language Technology for Cultural Heritage Data (LaTeCH)*, ed. Caroline Sporleder and Kiril Ribarov (Marrakech, Morocco, June 2008), 27–34.

[16] Marco Carlo Passarotti, "Development and Perspectives of the Latin Morphological Analyser LEMLAT," in *Digital Technology and Philological Disciplines*, ed. Andrea Bozzi, Laura Cignoni, and Jean-Louis Lebrave, vol. 20–1, Linguistica computazionale (Pisa: Istituti Editoriali e Poligrafici Internazionali, 2004), 397–414.

[17] Alexander Mehler et al., "Towards a Network Model of the Coreness of Texts: An Experiment in Classifying Latin Texts using the TTLab Latin Tagger," in *Text Mining. From Ontology Learning to Automated Text Processing Applications*, ed. Chris Biemann and Alexander Mehler, Theory and Applications of Natural Language Processing (Springer, 2014), 87–112.

ing them, and where I describe my own attempts of working on a tagger-lemmatizer for medieval Latin with Kestemont.[18]

The idea that additional normalization and morpho-syntactic information comes with advantages for stylistic analysis has often been proposed. Intuitively this makes sense: the better a model is equipped to capture the complexity of language, the better it should be able to understand a text's dynamics. In this light, one will often find a distinction in the literature between so-called 'shallow features,' which are the formal items that can readily be derived from the data as is, and 'deep features,' which require additional, more complex preprocessing. The promising prospects of deep features also seem to make sense when considering the success of function words in attribution cases. Function words are grammaticalized and often carry a syntactic rather than a semantic function. Therefore, a method taking into account the part of speech or syntactic functions of *all* tokens seems promising from a qualitative point of view. Lemmatization would likewise entail a more fine-grained selection of features, where the neighbouring context (and topic) of words is eliminated. Determiners, for instance, such as demonstratives (e.g. *hic, iste, ille*), possessive determiners (e.g. *meus, tuus*), quantifiers (e.g. *multus, plures*), distributive determiners (e.g. *quisque, quisquam*) and interrogative determiners (e.g. *quis?*) will be inflected in accordance with their associated referent. Surely, lemmatizing them to their headword will exclude the context-determined bias of gender and person in their frequency and statistical representativeness. However strong that intuition may be, it so turns out, however, that lemmatization and more complex annotation is not necessarily favourable for attribution accuracy. Lexical items still prove to be more stable and reliable. On the other hand, this moderate rate of success might also well be accredited to the fact that much depends not solely on the margin of error of the attribution method itself, but also on "the introduction of errors in the processing" when additional automatic tools are applied.[19] We will return to this much-debated topic in part 2 of this chapter, which provides illustrative experimentation.

## A.3.3 Sampling

### Segmentation into Text Samples

A final and fundamental preprocessing step is to decide upon sampling. At this stage, we have not yet covered the type of techniques that yield such accuracy percentages. We will return to this particular problem in part 2.

[18] Kestemont and De Gussem, "Integrated Sequence Tagging for Medieval Latin Using Deep Representation Learning," 3–6.

[19] Juola, "Authorship Attribution," 265.

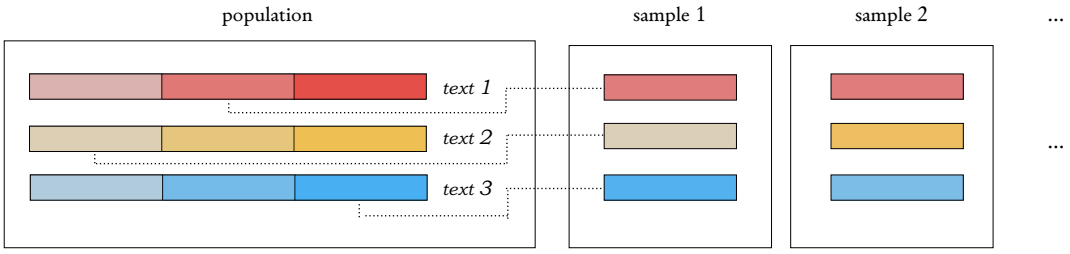



Figure A.1: Intuition of random sampling.

### Random sampling

Text sampling in the sense of segmentation should not be confused with another statistical technique, often simply denoted 'sampling,' which has a different intuition, and coincides with random sampling. In random sampling, some samples or subsets (sometimes called 'individuals') are randomly (i.e. by chance) drawn from a population in $n$ iterative turns (or in just a single turn, dependent on the experiment at hand). By no means such random sampling is exhaustive, the process does not imply that in the end the entire population will have been analyzed. The technique can be applied, for instance, if the population is simply too large to be statistically described in its totality (e.g. the well-known example of political surveys —survey sampling— before votings are cast). For computational stylistics specifically, such random sampling can be applied in cross-validation techniques, which will be discussed in A.3.5 (p. 340). There are different ways of random sampling. The variety of techniques originates from the idea that there are different ways of dividing up the original data set, and different ways of maintaining that original set's structure. Stratified random sampling, for instance, will not simply randomly sample from a population (this would entail that each individual has an equal probability of being selected), but will take into account the proportion of the original data set. Another interesting application of randomization, is one in which authentic sentences or authentic words of an author are selected and placed in random order to filter out the interference of atypical passages in the text, or passages where genre and register have taken the upper hand. The presence of outliers will be smoothened out towards a more uniform stylistic signal. This can also be a particularly interesting technique in the context of data augmentation, which is a specific scenario in which an analyst might feel that the original population is too limited in size (data sparsity): in consequence, the original population is artificially increased in size by imitating patterns within the original population.

### Data Augmentation and Generating Data

Instead of randomly sampling existent sentences from existent data, new language generation techniques for the sake of data augmentation have increasingly been proposed. Natural language generation (NLG) is a subfield of NLP, "defined as the task of gener-

ating text or speech from non-linguistic input."[20] Since the rapidly upcoming interest for the AI technology of deep neural networks (computing systems inspired by the human brain)[21] the domain of NLG is gaining significant traction. The idea is that a computer is trained to automatically reproduce language after having 'seen' a number of input examples on which it should base its output. The above is typical of a set of computational techniques collected under the denominator machine learning. We will treat language modelling and machine-learning techniques more fully at a later stage of this chapter (A.3.5, p. 322). It suffices to say, for now, that more complex generations of language appear to be favourable over simple, probabilistic reproductions of existent language. With Manjavacas, Daelemans and Kestemont, I have explored the potential for generated medieval Latin in authorship attribution tasks, on the basis of the *Patrologia Latina*.[22]

## A.3.4 Vectorization and Feature Extraction

### Feature Scaling

Since computational stylistics is oriented towards solving a problem of correlations, it is unwise to be working on raw feature frequencies, which is why they should be scaled. Feature scaling can also often be encountered in the literature under the denominator 'normalization' (not to be confused with token normalization in A.3.2, p. 314), or 'weighting.' Renunciation of this important step (which ensues after feature extraction) can have disastrous results, as was, for instance, one of Karian's main cases against the invalidity of the *Qsum* method discussed in some more detail in chapter 2 on p. 87. Measurements had been scaled differently in order to skew results.[23]

In fig. A.2, one gains a closer intuition. The tables in the figure list some hypothetical measurements, which are consequently scaled according to various methods. All of these methods are valid, although they might output some slight differences and some might perform better under certain circumstances than others. The following list gives a concise summary of three scaling methods that will often be applied in this thesis:

- **Min-max normalization**: all feature values are rescaled and gauged within the range of 0 (minimum value) to 1 (maximum value).
- **Standard scaling**: all feature values are standardized by subtracting the mean and scaling to unit variance (the mean for each word is 0, and the standard

[20] Albert Gatt and Emiel Krahmer, "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation," *Journal of Artificial Intelligence Research* 61 (2018): 65.

[21] Vivienne Sze et al., "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," *Proceedings of the IEEE (Institute of Electrical and Electronics Engineers)* 105, no. 12 (2017): 2295–2329.

[22] Enrique Manjavacas et al., "Assessing the Stylistic Properties of Neurally Generated Text in Authorship Attribution," in *Proceedings of the Workshop on Stylistic Variation* (Copenhagen, Denmark, September 2017), 116–25.

[23] Karian, "Authors of the Mind: Some Notes on the QSum Attribution Theory," 270ff.

| | Raw frequencies | | | Min-Max scaling | | | Standard Scaling | | | Tfidf-weighting | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *et* | *in* | *non* | *et* | *in* | *non* | *et* | *in* | *non* | *et* | *in* | *non* |
| document 1 | 89 | 46 | 32 | 0.64 | 0. | 1. | 0.34 | -1.68 | 1.60 | 89 | 46 | 39.14 |
| document 2 | 70 | 61 | 12 | 0.42 | 0.75 | 0.38 | -0.26 | 0.19 | -0.11 | 70 | 61 | 14.68 |
| document 3 | 34 | 65 | 0 | 0. | 0.95 | 0. | -1.42 | 0.69 | -1.13 | 34 | 65 | 0. |
| document 4 | 120 | 66 | 9 | 1. | 1. | 0.28 | 1.34 | 0.81 | -0.36 | 120 | 66 | 11.01 |

Figure A.2: Scaling methods.

deviation is 1). This is similar to z-score weighting, which we will discuss more elaborately on p. 331.

- **Tfidf weighting**: Tfidf stands for 'term frequency inverse document frequency.' It divides all feature values by the number of documents that respective feature appears in. As a consequence, less common features receive a higher weight, which prevents them from sinking away (and losing statistical significance) in between more common features.[24]

### Feature Selection

Feature selection is sometimes mixed up with feature extraction. Yet, the two processes have quite different goals and depend on differing techniques. Feature extraction can arguably be thought of as a first step, which is deriving values from the input data (a corpus of texts). For instance, extracting the 150 most common words as features on which the subsequent vectorization takes place, as we have done a few pages earlier, can be understood as feature extraction. Feature selection, on the other hand, is a second step. It aims "to select a subset of variables from the input which can efficiently describe the input data while reducing effects from noise or irrelevant variables and still provide good prediction results."[25] In other words, it chooses the best possible result with the least possible means. With feature selection, an algorithm is activated to filter out redundant features from the dataset, and retain features that yield conspicuous differences (or variations) in frequency: after all, those will often contain the most information. The advantage of feature selection is that it simplifies interpretation; it avoids distraction from those features that are less relevant and cause unwanted additional noise in the quest for detecting difference between texts. Since feature selection also decreases the length of the list of features, it can simultaneously be thought of as a dimensionality reduction method. In scholarship, one will encounter that feature selection methods are often categorized as being either a 'filter method' or a 'wrapper

[24] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval* (New York: Cambridge University Press, 2008), 109.

[25] Girish Chandrashekar and Ferat Sahin, "A Survey on Feature Selection Methods," *Computers and Electrical Engineering* 40 (2014): 16.

method,' which in some cases are combined to an 'embedded method.' In order to discuss these feature selection methods in full, however, some prior knowledge on machine learning and classification is advised, which is why we will return to this division only in second part of this chapter.

## A.3.5 Classification

### Machine Learning and Classification

Machine learning is a field in computer sciences that has in the past decades increasingly gained impact due to the quickly developing technical achievements in computing since the 1980-90s.[26] Machine learning techniques have likewise proven to be "an important turning point" in the field of computational authorship attribution.[27] Multiple reasons lie at the basis of why machine learning has been revolutionary for attribution studies and natural language processing in general.[28] One is that machine learning classifiers allow to digest larger data sets with longer feature vectors, enabling scholars to better assess the accuracy and reliability of stylometric methods, e.g. identifying a writer in a set of no less than 145 authors.[29] Classification methods have also allowed to systematically and exhaustively experiment with and assess series of divergent input vector representations and parameters. Such practices can occasionally contribute to the overall reputation and validity of computational attribution, and presents one with an optimistic outlook of the field. At other times, such large-scale data-driven stylometric experiments invoke some despondency. For medieval Latin specifically, Maciej Eder has shown recently —in an enormous *Patrologia Latina* benchmark study of 5,281 texts that took weeks to run— that authorship attribution scores with "the optimal attributive success achieved [were] as high as 50% for the most effective set of input parameters."[30] In other words: in the greater scheme of things, if one seeks to attribute an anonymous medieval Latin text that is possibly written by one of approx. 5,300 Church Fathers, in a best case scenario our reliability threshold is set at a 50%

[26] On the effect of machine learning on authorship attribution in the 1990s, see for instance Stamatatos, "A Survey of Modern Authorship Attribution Methods," 539. The bibliography on machine learning is immense and ever increasing. For a recent introduction to machine learning, now at its second edition, see Xia Jiang and Richard E. Neapolitan, *Artificial Intelligence: With an Introduction to Machine Learning*, 2nd ed. (Boca Raton: CRC Press, 2018).

[27] Moshe Koppel, Jonathan Schler, and Shlomo Argamon, "Computational Methods in Authorship Attribution," *Journal of the American Society for Information Science and Technology* 60, no. 1 (2009): 11.

[28] According to Jurafsky and Martin, three synergistic trends can be gathered. The first is the boom of electronic disclosure and availability of data, the second is the ensuing interplay between machine-learning and linguistics, and the third is the democratization of high-performance computing systems, see Dan Jurafsky and James H. Martin, *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd ed., Prentice Hall Series in Artificial Intelligence (Upper Saddle River, New Jersey: Pearson Education International, 2009 (2000)), 46–7.

[29] Kim Luyckx and Walter Daelemans, "Authorship Attribution and Verification with Many Authors and Limited Data," in *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)* (Manchester, UK, August 2008), 513–20.

[30] Eder, "Taking Stylometry to the Limits."

chance that the output label is correct.

### Intuition of Classification

Consider the following example: in authorship attribution, we expect the computer to observe text $x$ in order to recognize that it was written by author $y$. However, reliable prognosis requires some training. Think of our algorithm as an ornithologist who can only name a bird once he or she has gained expertise in distinguishing and recognizing all possible birds on the basis of their distinctive features. The computer likewise needs to see real-life examples from which it can learn to know a bird by its feathers. Before it can acquire this skill, the computer is fed with data in a paired $< x, y >$ format, which corresponds to <text input, author label>, or <input vector $\vec{X}$, output label $Y >$ (cfr. fig. 3.4). We have seen earlier that our input vector $\vec{X}$ is in fact a numerical representation of a text fragment, i.e. an abstraction of the original text in terms of its most conspicuous features. Vector $\vec{X}$ packs together an extensive amount of information which can now be taught to stand in correlation with a specific label or class. By carefully examining this training data, an algorithm is able to draw distinctions between the samples on the basis of such numerical trends. Once the algorithm has learned to distinguish between reference samples in the data, a user expects the algorithm to be capable of making well-informed predictions on test data, i.e. unseen samples, texts of unknown authorship, whose suspected author is a member of the potential classes.

Intuitively this all sounds relatively simple, but there are a few pitfalls attached to the learning process of a classifier. In fact, a recurrent challenge is to get a grip on how reliable the model is, and how to evaluate the decisions that it makes. This is because there are in fact many different classifier models (to name but a few: support vector machines, *k*-Nearest Neighbour classifiers, nearest shrunken centroid classifiers, Naive Bayes, neural networks, ...), which all react differently to different problems, have a variety of parametrization options and require other methods by which to optimize their performance during training. Ultimately, even the stance from which we evaluate (or score) a classifier as fit or unfit for solving the classification problem at hand, is subject to discussion, and dependent on many localities. We will see, moreover, that the evaluation scores (percentages of how 'accurate' a model is) of their decision-making are always to be taken with a grain of salt, and by no means guarantee an absolute certainty (cfr. A.3.5). This does not, however, mean that there are not some preparatory measures one can take in order to bypass the unpredictability of classifiers, or should give up to aim for a robust model which should avoid as much uncertainty as is statistically feasible.

In the example in fig. A.3, the anonymous snippet $\vec{X}$ is likely to have been written by Bernard of Clairvaux, because its coordinates make it fall on "Bernard's side" of the decision boundary. It is the latter equation (the decision function) by which

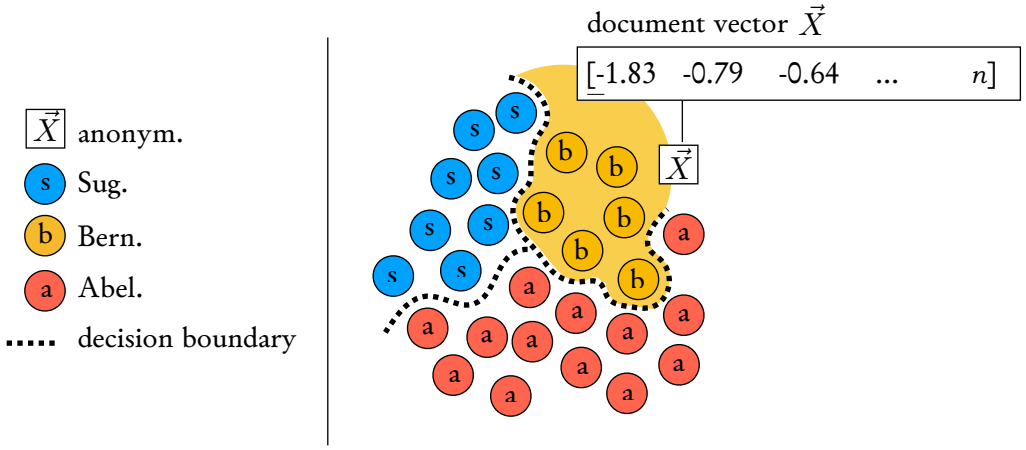



Figure A.3: Intuition of decision boundary drawn by classification algorithm in between a hypothetical set of document vectors appertaining to Suger of Saint-Denis (blue), Bernard of Clairvaux (red) and Peter Abelard (yellow) to demarcate classes. This figure is a rehearsal of fig. 3.5 on p. 96.

classifiers 'draw' a decision boundary. This equation can cause the classification on the one or other side of the margin to differ. As we mentioned, this means that one can apply different classifiers to the same problem, yet might ultimately end up with different margins and therefore different predictions for the test data. Each classifier has its own unique way of recognizing patterns in data (classification is considered a subfield of 'pattern recognition'), and it is often impossible to know in advance which classification model will suit the problem at hand best.

In what follows, we will explore some of the aforementioned types of classifiers. The aim is not to exhaustively survey all classifiers and their specificities. I will choose my examples by virtue of their being fit to illustrate particularly well the basic mechanisms that are universal to all algorithmic classification models. The main aim is to provide only an introduction to the concept of classification, by gently sliding from a probabilistic, mathematical intuition of classification into one that is geometric and visual. Not coincidentally, the reader more experienced in classification models will notice that this implies a shift from so-called 'probabilistic' (A.3.5) models to 'similarity-based' (A.3.5) and 'vector space' (A.3.5) models.[31] Throughout this dissertation, the specificities and parameters of different classifiers that some readers might still find lacking in this chapter will be documented more fully as we proceed with more real-time experiments.

### Probability Theory and Naive Bayes

It is beyond the scope of this thesis to explore the field of probability theory in full. This section scratches the surface of the concept, since it is vital to raise the awareness that probability theory is the underlying logic behind all of the classification experi-

[31] I am borrowing this taxonomy of classifiers from Stamatatos, "A Survey of Modern Authorship Attribution Methods," 548–9.

ments in this dissertation. In more concrete terms, and rephrased from the current perspective of computational authorship attribution for medieval Latin, we eventually ask little more than: given a medieval Latin text of unknown authorship, what is the 'probability' that this text was written by the author(s) that I have in mind? Here we will take as example an accessible and basic probabilistic classifier,[32] which serves as a good stepping stone to initiate the probabilistic operations behind classification: Naive Bayes.[33] This is not a coincidental choice. Frederick Mosteller and David L. Wallace, the pioneers of computational authorship attribution, used Bayesian methods in their landmark study on the Federalist Papers.[34] Ever since, the Naive Bayes classifier has been revisited frequently, labelled —not always deservingly— as a bad estimator for classification in authorship attribution.[35] Indeed, in the meantime, its workings have been overtaken by other more sophisticated machine-learning models, yet Naive Bayes —in all its charmingly honest naivité— still often proves to be a useful benchmark model for testing other classifiers.[36] The model moreover still serves as an ideal starting point to illustrate the mathematical principles behind classification for the uninitiated.

A probabilistic classifier such as Naive Bayes (cfr. fig. A.4.)[37] is a type of classifier that computes and outputs probabilities for each class $y$ under consideration by calculating the probability of each and every feature $x$ in the training dataset. In the problem where we have our three hypothetical classes named Suger, Bernard and Abelard, a Naive Bayes algorithm would present us with the respective odds (a percentage) that an unknown document vector $\vec{X}$ was written for each of these hypothetical classes. Naive (or 'simple') Bayes departs from a probability theorem that is —as its name suggests— somewhat naive: during training it assumes independency between all the features (cfr. the Bayesian equation on the right-hand side of fig. A.4), and considers all features to contribute equally to the learned label output $y$ in its decision equation.

The reason why such an independence assumption between features is considered 'naive,' is because one cannot reasonably expect that features function independently in a sort of vacuum, uninformed by the existence of other features. Let us revisit the example of the ornithologist. Say, for instance, that (s)he wants to classify some unknown bird. Amongst many features, the bird scores high on such features as

[32] For a detailed study on probabilistic classifiers, see Fabrizio Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys* 34, no. 1 (2002): 1–47.

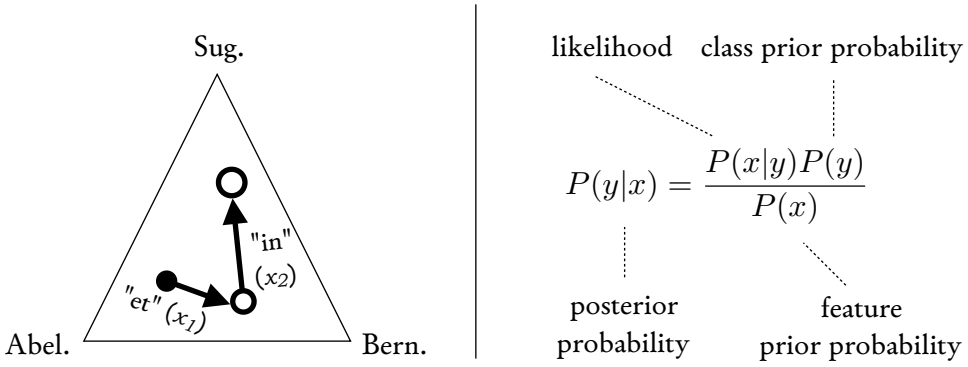
[33] Geoffrey I. Webb, "Naïve Bayes," in *Encyclopedia of Machine Learning*, ed. Geoffrey I. Webb and Claude Sammut (New York: Springer, 2011), 713–4.

[34] Mosteller and Wallace, *Applied Bayesian*.

[35] Another name in sway is "Idiot's Bayes." See Hastie, Tibshirani, and Friedman, *Elements*, 210.

[36] See for instance Bei Yu, "An Evaluation of Text Classification Methods for Literary Study," *Literary and Linguistic Computing* 23, no. 3 (2008): where Yu demonstrates that although Naive Bayes is often outperformed by support vector machines, it is not necessarily in all text classification scenarios the better model.

[37] This diagram is largely inspired by a figure in Bird, Klein, and Loper, *Natural Language Processing with Python*, 246. It has, however, been tailored to illustrate Naive Bayes for an authorship attribution problem instead of an information retrieval or document classification problem.




$$P(y|\vec{X}) = P(x_1|y) \cdot P(x_2|y) \cdot ... \cdot P(x_n|y) \cdot P(y)$$

Figure A.4: Intuitive diagram of Naive Bayes on the left. In a case where Peter Abelard has written the most documents (13 texts) in the training set, the classifier is automatically poised towards that label in an initial position. Then the decision is redirected by how each feature $x$ influences the probability outputs. The right-hand side features Bayes' theorem. The theorem outputs posterior probabilities for each of the features, which are consequently multiplied (cfr. the bottom formula).

'rearrangeable feathers' ($x_1$), 'lightweight' ($x_2$), 'aerodynamic' ($x_3$) and 'fast' ($x_4$), making the bird spotter conclude that (s)he is observing a bald eagle ($y$). Although this classification might perfectly be correct, features such as 'lightweight,' 'aerodynamic' and 'rearrangeable feathers' all contribute to and correlate with the feature of 'speed.' We are, therefore, quadrupling the importance of one specific aspect of the bird — its speed— aggregating the respective posterior probability scores for these features. Since Naive Bayes treats every feature independently, it does not take correlation into account, and therefore runs the risk of exaggerating one and the same observation, causing it to become overconfident in decision-making. Another typicality of Naive Bayes is that for its predictions it takes into account a normalizing 'prior probability' of the output label $y$ before coming to a 'posterior' conclusion ('posterior probability'). In the ornithologist metaphor, the implication of such a prior probability is that rarer birds will more unlikely be predicted, as the chances of coming across them are already substantially smaller.[38]

Within a computational authorship attribution context, a Naive Bayes model 'learns' as follows: let us take as example corpus the hypothetical corpus which was depicted in fig. A.3 on p. 324, and consisted of 8 texts written by Suger of Saint-Denis, 6 by Bernard of Clairvaux, and 13 by Peter Abelard. Let us assume that these texts are all samples or instances of a class consisting of 1,000 words (we discussed segmenta-

[38] Note that in authorship attribution, this type of weighting according to class will rarely be useful, as it in fact induces class imbalance problems. We would rather take any author equally seriously into account, regardless of whether or not the respective author was prolific. On the other hand, one could argue that less prolific or less known authors are —in terms of probability— less likely to have written a transmitted document. Transmission is, of course, rather slippery territory in medieval studies.

| | | Labels $y$ | | |
|---|---|---|---|---|
| | | Sug. (8 docs.) | Bern. (6 docs.) | Abel. (13 docs.) |
| Features $x$ | 'et' ($x_1$) | $\frac{401}{8000}$ | $\frac{621}{6000}$ | $\frac{619}{13000}$ |
| | 'in' ($x_2$) | $\frac{511}{8000}$ | $\frac{331}{6000}$ | $\frac{498}{13000}$ |
| | 'est' ($x_3$) | $\frac{61}{8000}$ | $\frac{82}{6000}$ | $\frac{137}{13000}$ |

Table A.3.1: Hypothetical table of likelihoods —$P(x|y)$— for each respective class.

| | | Labels $y$ | | |
|---|---|---|---|---|
| | | Sug. (8 docs.) | Bern. (6 docs.) | Abel. (13 docs.) |
| Posterior probabilities | $P(y|`et')$ | 0.244 | 0.378 | 0.377 |
| | $P(y|`in')$ | 0.381 | 0.247 | 0.372 |
| | $P(y|`est')$ | 0.218 | 0.293 | 0.489 |

Table A.3.2: This table provides the computed posterior probabilities —$P(y|x)$— for each of the features $< x_1 ... x_n >$. Note that adding up the probabilities in the rows amounts to 1.00.

tion and sampling in chapter A.3.3). This would mean that in total our hypothetical training corpus holds 8,000 words ($w$) by Suger of Saint-Denis, 6,000 $w$ by Bernard of Clairvaux and 13,000 $w$ by Peter Abelard. A Naive Bayes probabilistic model observes and computes for each feature in the training data the probability that this feature was used by an author, whilst taking into account the other authors in the training set. In this sense, Naive Bayes constructs a chain of probability outputs, which can ultimately be multiplied towards a total posterior probability. The result of this multiplication is the posterior probability, i.e. $P(y|x)$ or in our case $P(Suger|`et')$.

Consider fig. A.4, which provides an abstract illustration of how Naive Bayes chains up probabilities on the left, and the mathematical formulation of the theorem on the right. In addition, consider table A.3.1. Firstly, Naive Bayes calculates the **likelihood** by which the word 'et' can be encountered in Suger's corpus:

$$P(x|y) = P(`et'|Suger)$$

Explicitly, this means the following: if we have 8,000 words by Suger in total, and 401 of them are the conjunction 'et,' the model learns that the likelihood of finding 'et' in a text by Suger is $\frac{401}{8000}$ or 0.050 (cfr. table A.3.1).

The likelihood is consequently multiplied by the odds that we will come across a text by Suger in general, the **class prior probability:**

$$P(y) = P(Suger)$$

In this specific situation, where we can speak of 27 individual text samples of which 8 are by Suger, the prior probability of labelling a text as by "Suger" as such is $\frac{8}{27}$ or 0.296. One can think of prior class probability as encoding an observer's intuition beforehand, i.e. weighting and qualifying one candidate as more likely than the other.

Consequently, this score above the denominator is then divided by another prior probability, namely the probability that 'et' will occur throughout the entire corpus. This is the **feature prior probability**, which can be achieved by adding up all the instances of 'et' and comparing this to the total number of words:

$$P(x) = P(`et')$$

The feature prior probability in this example equals $\frac{1641}{27000}$ or 0.061. Note that this value is a constant that mainly serves to normalize the equation, and is not always taken into consideration.

If we insert the likelihood, class prior probability and feature prior probability into the Bayesian theorem as shown on the right-hand side of fig. A.4, we arrive at table A.3.1 with the **posterior probability output:**

$$P(y|x) = P(Suger|`et')$$

The result of the posterior probability in 'et,' just as some example posterior probabilities of other words, can be discovered in table A.3.2 to amount to 0.244. If one looks closely, one will see that the values in this table correspond to the decision movements the algorithm makes in the diagram on the right-hand side of fig. A.4. In this hypothetical situation, the odds are higher to find a lot of 'et' conjunctions in Bernard of Clairvaux, and the odds are higher to find the preposition 'in' with a higher frequency in the corpus of Suger of Saint-Denis.

The class-specific posterior probabilities as tabularized in table A.3.2 furnish the evidence / empirical data that allows a Bayesian model to infer decision rules, i.e. to generalize what it has learned (hence, *Bayesian inference*). It does so by assuming that the observations seen in the training data are but a fraction of a larger, unknown — and therefore 'posterior'— population that follows a type of probability distribution. There are multiple types of distributions imaginable,[39] but for clarity's sake one could best imagine a continuous probability distribution as the common curve-shaped normal distribution (also known as a Gaussian distribution or 'bell curve,' see fig. A.5). The latter implies that on the basis of the behaviour of 'seen' documents within a class, the model expects that future 'unseen' documents within that same class will behave according to this distribution, where most measurements will float around a learned mean ($\mu$), and deviate according to a learned standard deviation ($\sigma$). This fitting of a

[39] E.g. the multinomial, Bernoulli or out-of-core distributions, see Susana Eyheramendy, David D. Lewis, and David Madigan, "On the Naive Bayes Model for Text Categorization," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistic*s, ed. Christopher M. Bishop and Brendan J. Frey (Key West, January 2003).

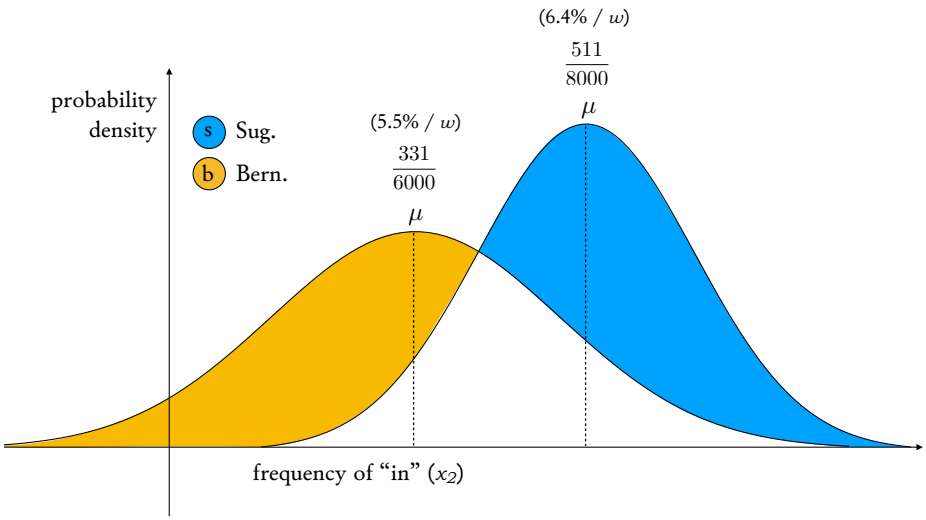



Figure A.5: Intuition of Gaussian posterior probability distributions fitted as training data to MAP. The multiplication of these prior distributions (the evidence) is what will ultimately yield a posterior probability distribution. Given a document where the word 'in' is good for approximately 5.5% of that document's vocabulary, then that document will fall within the yellow region of Bernard of Clairvaux, in other words the region where $\frac{331}{6000}$ presents the mean. If 'in' is good for 6.4% of the vocabulary of a document, then it falls within the blue region appertaining to Suger of Saint-Denis (where the mean is $\frac{511}{8000}$).

distribution to data, taking the mean $\mu$ and the standard deviation $\sigma$ as parameters, is called 'maximum a posteriori estimation' (MAP).

MAP's fitting of a distribution is a way for the model to store and memorize information, i.e. to summarize what it has learned from 'prior' evidence by defining and remembering the parameters which describe a 'posterior' distribution. The intuition behind MAP is in fact simple: we assume that the statistical model that best describes the evidence will also be the model that will best describe future observations. In other words: we estimate the parameters that describe our posterior probability best, and *maximize* the likelihood that future data will be well classified:[40]

$$y_{\text{MAP}} = \underset{y}{\operatorname{argmax}}\, P(x|y) \times P(y) = \underset{y}{\operatorname{argmax}}\, P(y|x)$$

The gist of classification, then, here understood from the perspective of Naive Bayes, is that it attempts to capture a set of probability distribution parameters that make a good fit (*optimize*) to the prior distributions which were based on the evidence. Doing this for a single feature is intuitive enough and can be visualized quite easily. However, for a model to perform better, we want to take into account not solely the distribution of 'in' as exemplified in figure A.5. MAP learns the distributions (the mean $\mu$ and the standard deviation $\sigma$) and parameters $\theta$ of many individual features: 'et,' 'in,' but also

[40] In that sense, MAP is much akin to maximum likelihood estimation (MLE), except that MLE does not take into account a prior distribution (i.e. prior probability —$P(y)$— in the distribution fit. Instead, MLE departs from what is a 'uniform prior.'

others. As mentioned in the paragraph on feature independence above (p. 325), Naive Bayes assumes that each feature has its own normal (or some other) distribution, which equally weighs in on the ultimate label decision. In order to take into account a larger vocabulary, the respective posterior probabilities are to be multiplied:

$$P(y|\vec{X}) = P(x_1|y) \times P(x_2|y) \times ... \times P(x_n|y) \times P(y)$$

The multiplication of all these features —which usually happens in log space— will ultimately present us with a set consisting of *n* parameters $\hat{\theta}$ that describe the reality of our given data best (the hat on top of the letter denotes that $\theta$ is considered an estimate output). Therefore, we can rephrase Bayes' theorem in a way that it represents our classifier's higher-order goal which is no longer solely to maximize the likelihood of correctly classifying a given label by looking at a single given feature, but in fact to maximize the likelihood of correctly classifying *all* labels by looking at *all* features:

$$\hat{\theta} = \theta_{\text{MAP}} = \underset{\theta}{\operatorname{argmax}}\, P(data|\theta) \times P(\theta) = \underset{\theta}{\operatorname{argmax}}\, P(\theta|data)$$

We want to maximize the likelihood that our estimate $\hat{\theta}$ describes our original data: $\hat{\theta}$, the full set of parameters to our model, is optimized towards an as accurate as possible numerical description of our data.

Although the previous section might seem somewhat dense and field-specific, it mainly serves to arm us with a few key insights. One such insight is that, in classification, the differences amongst texts —and in their own turn, the differences amongst authors— can be mapped as continuous distributions; values that are in their own turn treated as if they were —in statistical terms— a population or sample space with a mean ($\mu$) and a standard deviation ($\sigma$), from which probability rules can be inferred. Another insight is that these probability distributions can be numerically captured through parameters. In the example above this role was taken up by $\theta$. Optimizing $\theta$ is consequently a classifier's main goal. After all, an optimal interpretation of $\theta$ will yield an optimal decision function, and an as accurate as possible model for future predictions.

### Burrows's Delta

As already hinted at earlier, the decision function of a classifier is based upon probability theory, but it is generally easier to wrap one's head around classification by visualizing its actions from a geometric perspective. Whereas in the previous section we have mainly been describing the differences between texts as 'mathematical differences,' these differences can also be interpreted as distances. Similarity is measured by how close one vector is to another, and a classifier is trained to map out these vector's locations (coordinates) in the training data. For exploring the geometrical intuitions of classification, we shift from the Naive Bayes classification model to Burrows's Delta, an extremely popular and widely used means of calculating the distances

between samples. Delta provides an ideal means of gently sliding from a probability-based explanation of classification to one that is geometrical.

### From Probability to Distance

Inspired by Mosteller and Wallace's observation that Naive Bayes can classify texts according to author on the basis of probability distributions, John F. Burrows in the beginning of the 21st century introduced a measure of difference that would break ground for the field of stylometry. The measure of difference was termed "Burrows's Delta,"[41] and although it was based on many of the same statistical principles already explored by Mosteller and Wallace, it was innovative in terms of its simple effectiveness for authorship attribution. His Delta ($\Delta$), which —as we will show below— was "the mean of the absolute differences between the z-scores for a set of word-variables in a given text-group and the z-scores for the same set of word-variables in a target text,"[42] proved to give highly accurate results for authorship attribution. Moreover, Burrows's method tackled what is —and will perhaps always remain— a pickle in authorship attribution: the 'open game,' and by extension *the needle-in-a-haystack problem*.[43] What if many authors are candidates that should be seriously considered? How can one be certain that attributions based on probability distributions are not in some way coincidental? Burrows consequently introduced the concept of including a much larger main dataset of potential authors in the experiments, and —much akin to the Naive Bayes principles explained above— approached this larger set of training data as statistical data to which a probability distribution can be fitted (norms and deviations; means and outliers). Ultimately, each author will have his/her place in this distribution, enabling his/her own stylistic identity to be classified.

In fig. A.6 above, an intuition of how to think of this training set distribution is depicted. Imagine, for instance, that this specific distribution curve was fitted on the relative frequencies of —again—the Latin preposition 'in' . The mean ($\mu$) on top of the distribution denotes the norm throughout this training set. The respective standard deviations ($\sigma$) on either side of the mean indicate the amount of expected deviation for the usage of the word 'in' from this mean. Consequently, authors can be ordered on this curve from the viewpoint of "deviation from a norm."[44] A common way to quantify this in statistics is z-scores. z-score is little else than the number of standard deviations away from the mean (see the top right corner of fig. A.6). Burrows advocated the use of z-scores because they ensure that for each word in our feature vector the frequencies will be normalized within a comparable range, or in other words, less frequent words' frequencies will not lose their impact in between

[41] Burrows, "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship."

[42] Ibid., 271.

[43] Tempestt et al., "Surveying Stylometry Techniques and Applications," 86:2.

[44] Stefan Evert et al., "Understanding and Explaining Delta Measures for Authorship Attribution," *Digital Scholarship in the Humanities* 32, no. suppl_2 (2017): ii5.

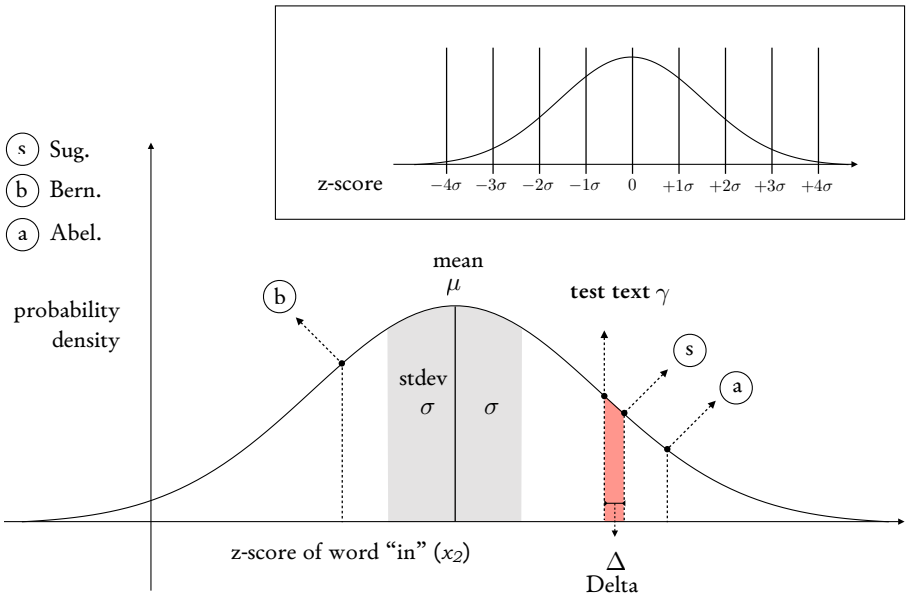



Figure A.6: Intuition of Burrows's Delta ($\Delta$).

the scores of highly frequent words:

> The z-scores are used to obtain cognate figures for all the words in a hierarchy where the original frequencies fall away sharply from top to bottom. The object is to treat all of these words as markers of potentially equal power in highlighting the differences between one style and another.[45]

Consequently, Burrows proposes Delta as a method that seeks an "expression of difference, pure difference."[46] The idea is that when we have some test text $\gamma$, we have to determine where the text can be fitted on the distribution curve, in order to know its respective distance from all of the authors in the training set. In fig. A.6, the delta score from Suger compared to test text $\gamma$ is indicated in red, because it highlights the smallest absolute difference between z-scores, i.e. smallest delta score, i.e. smallest distance. If we were to perform this calculation for each and every feature (not just 'in' ($x_2$) but many others), and if it turned out that delta score $\Delta$ between test text $\gamma$ and Suger was consistently low for many other features, the mean delta score over all features would point out that —amongst many other authors— Suger was most likely the author of test text $\gamma$ of all the authors in the training set. Note that the mathematical reasoning of Delta —if one turns back a few pages— is in fact highly similar to Naive Bayes (A.3.5).

[45] Burrows, "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship," 271.
[46] Ibid., 269.

### Similarity Measures

What is interesting about Delta ($\Delta$), is that the metric proves to be an ideal stepping stone to take the probabilistic interpretation of classification towards its geometric interpretation. As Shlomo Argamon demonstrated some six years after Burrows's discovery, Delta is "a kind of 'distance measure,' where the 'nearest' authorship candidate is chosen for attribution."[47] So much could in fact already be gathered from fig. A.6 above. Sheerly by looking at distance on the distribution curve (in red), one could already visually deduce that Suger was the most likely author of target text $\gamma$. Argamon not only made Delta more comprehensible. He also showed that the Delta measure, although it was designed specifically for computational authorship attribution, is a variation upon old and simple statistical reasoning according to distributions, and is in fact highly similar to many other classification and clustering techniques. Argamon stated that "[...] Delta may be viewed as an axis-weighted form of 'nearest neighbor' classification [...], where a test document is classified the same as the known document at the smallest distance."[48] This in its own turn spawned reactions and adaptations (or even improvements) of the Delta measure by others,[49] which in the meantime has caused the algorithm to deal with some considerable competition.[50] The central idea which we need to hang on to, is that the pairwise differences between two texts can be interpreted as a 'geometric distance' "between representations of documents as vectors in a high-dimensional space in which each word (or other feature) taken into account corresponds to one of the dimensions of that space."[51] It so turns out, however, —and Argamon himself emphasized this— that a distance measure is not without some controversy.

### Euclidean, manhattan and cosine

There are quite a few distance measures (or distance functions) imaginable, but it is defendable to argue that these are all variations upon three commonly encountered measures, respectively called 'manhattan,' 'euclidean' and 'cosine' distances (cfr. fig. A.7).[52] 'Euclidean' distance is commonly the most straightforward way of going from one place to another, and can be generally thought of as the distance in which

[47] Shlomo Argamon, "Interpreting Burrows's Delta: Geometric and Probabilistic Foundations," *Literary and Linguistic Computing* 23, no. 2 (2008): 132.

[48] Ibid.

[49] David L. Hoover, "Delta prime?," *Literary and Linguistic Computing* 19, no. 4 (2004): 477–95; Peter W.H. Smith and W. Aldridge, "Improving Authorship Attribution: Optimizing Burrows' Delta Method," *Journal of Quantitative Linguistics* 18, no. 1 (2011): 63–88; Eder, "Taking Stylometry to the Limits."

[50] Matthew L. Jockers and Daniela M. Witten, "A Comparative Study of Machine Learning Methods for Authorship Attribution," *Literary and Linguistic Computing* 25, no. 2 (2010): 215–23.

[51] Evert et al., "Understanding and Explaining Delta Measures," 3.

[52] The figure here is inspired to a very large extent on a figure in the important work of Stefan Evert et al., "Understanding and Explaining Delta Measures for Authorship Attribution," *Digital Scholarship in the Humanities* 32, no. suppl_2 (2017): and is here solely adapted to suit some of our current interests.

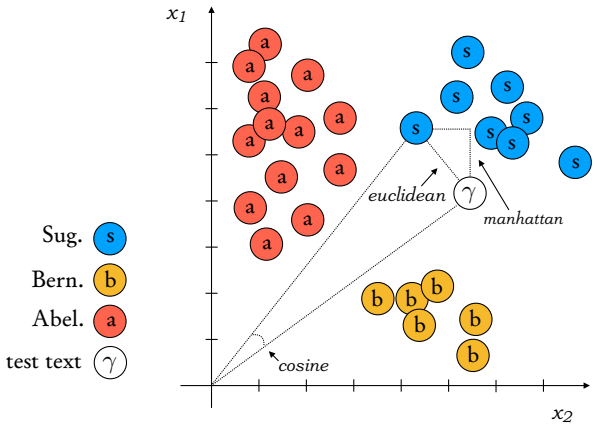



Figure A.7: Intuition of distance measures 'manhattan,' 'euclidean' and 'cosine' for test text $\gamma$ against one sample of Suger's writing.

'the crow flies,' directly fixing one point to another without taking any turns. The 'manhattan' distance or 'taxicab distance,' however, which has its onomasty in a city street plan, is a distance function which measures two points along axes at right angles, indeed, as if one would be navigating in a city along city blocks. Lastly, the 'cosine' similarity is perhaps the least intuitive, as instead it measures the cosine of the angle between the points.

As one can imagine, the route one chooses to take from origin to target text can considerably alter the resulting similarity measure. If one were to try and find the three closest neighbours to a point, it is possible that the three candidates are ranked differently or differ altogether, depending on the type of distance function one chooses. A possible solution, which we will return to later, is to find a 'consensus' amongst these ranked candidates by iterating a number of times along different parameters. This is the idea of bootstrapping.[53]

## Vector Space Classification

There are different ways of demarcating spaces and drawing a decision boundary (or —for multidimensional scenarios— decision hyperplanes) as we have visualized in fig. A.3 on p. 324. Classifiers that rely on such a partitioning of space are named vector space classification models. Such models rely on the so-called contiguity hypothesis, which means that "documents of the same class form a contiguous region and regions of different classes do not overlap."[54] What such vector space models have in common in their quest for distinctive regions, is that they set out to detect 'landmarks' in the

[53] See for instance Eder's work on consensus networks in Maciej Eder, "Visualization in Stylometry: Cluster Analysis Using Networks," *Digital Scholarship in the Humanities* 32, no. 1 (2017): 50–64; or his earlier work on bootstrapping methods for the Delta distance in Maciej Eder, "Bootstrapping Delta: a Safety Net in Open-Set Authorship Attribution," in *Digital Humanities 2013: Conference Abstracts* (University of Nebraska–Lincoln, July 2013).

[54] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval* (New York: Cambridge University Press, 2008), 289–90.

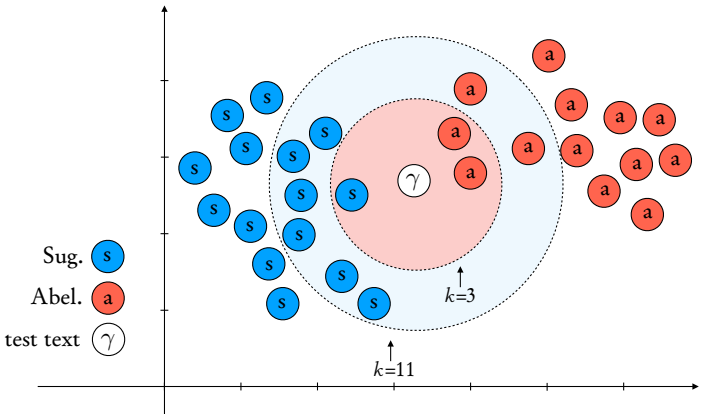



Figure A.8: Intuition of $k$ Nearest Neighbours model, exemplifying the importance of parameter $k$. If the number of neighbours $k$=3, test text $\gamma$ will be attributed to the class in red (Abelard). If the number of neighbours $k$=11, test text $\gamma$ will be estimated as a part of the blue class (Suger).

data which allow them to coordinate in a computationally less expensive way than having to compute every coordinate in that space.

The preceding can be rephrased as follows: think of an explorer who discovers a new land, and consequently wishes to navigate in that land by mapping out several distinct regions. It makes very little sense for that explorer to remember each and every distinct point in that region in order to find his/her way back home. What the explorer needs is an abstraction, a map, preferably one that is a generalization and not a detailed, lengthy description of the unexplored land. Instead of remembering each and every coordinate, (s)he can make use of landmarks, recognizable aspects of that region that are either 1) at the core of or 2) in the bordering, transitional zone of the region. In fact, the differences in approach available to the explorer —options 1 and 2 in the preceding sentence—, correspond in general terms to two commonly used approaches in vector space classification. Either one characterizes and learns the regions by the points that typify that region most and stand at the centre, or either one attempts to find the area of transition, the separating border between the regions. The first method roughly corresponds to a technique called $k$ Nearest Neighbours, and the other to support vector machine modelling (SVM).

### *k* Nearest Neighbours

In a $k$ Nearest Neighbours classification model, as illustrated in fig. A.8, our landmarks are the points closest to test text $\gamma$ (centrally in the figure). The class best represented in number —i.e. majority class— by these nearest neighbours determines the class of $\gamma$.[55] In other words: "we expect a test document $d$ to have the same label as the

[55] Pádraig Cunningham and Sarah Jane Delany, "$k$-Nearest Neighbour Classifiers," *Multiple Classifier Systems* 34 (2007): 1–17.

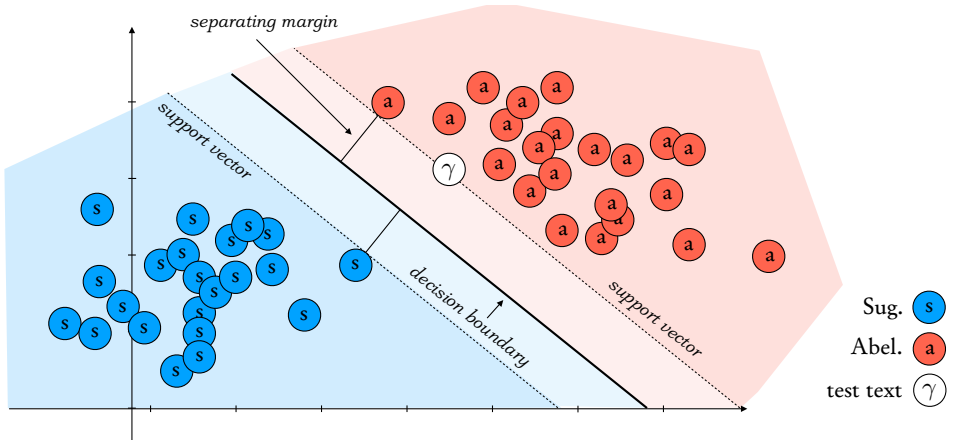



Figure A.9: Intuition of Support vector machine (SVM) for a two-class problem (linear kernel).

training documents located in the local region surrounding *d*."[56] It is not uncommon practice in some variations of the algorithm that neighbours lying closer (i.e. with a higher 'similarity measure,' cfr. subsection A.3.5 above) weigh in more heavily on the algorithm's decision-making.[57] On a sidenote: a commonly encountered variation upon the *k*-NN technique is the nearest centroid classifier or Rocchio classifier.[58] In this method, a mean for each of the classes is computed (synonymous with such terms as prototype or centroid). It is the distance between the test example $\gamma$ and a 'nearest mean' (instead of some nearest neighbour) that consequently lets the algorithm decide on how to classify the test example.[59]

### Support vector machines

Another popular vector space classification technique worth mentioning here —especially considering its wide use within authorship attribution problems— is SVM (support vector machine).[60] Much akin to *k* Nearest Neighbours, it constructs landmarks from seen training examples. Yet, instead of taking as reference the most neighbouring instances of a class or a prototypical numerical representation of that class (i.e. a centroid), SVM tracks down the instances of a class that lie in a gray zone, and uses those instances to span a boundary from one end to another. In this sense, it is a classifier that shows a different approach in how to organize its spatial demarcation than *k*-NN.

As seen in fig. A.9, SVM departs from the assumption that a two-class or multi-class

[56] Manning, Raghavan, and Schütze, *Introduction to Information Retrieval*, 273.

[57] Sahibsingh A. Dudani, "The Distance-Weighted *k*-Nearest-Neighbor Rule," *IEEE Transactions on Systems, Man, and Cybernetics* 6, no. 4 (1976): 325–7.

[58] Manning, Raghavan, and Schütze, *Introduction to Information Retrieval*, 269.

[59] It is in that sense the supervised alter ego of the unsupervised *k* means clustering technique.

[60] For a concise but insightful introduction to support vector machines, see the entry in the *Encyclopedia of Machine Learning*: Xinhua Zhang, "Support Vector Machines," in *Encyclopedia of Machine Learning*, ed. Claude Sammut and Geoffrey I. Webb (New York: Springer, 2011), 941–5; for SVM's specifically applied to computational authorship attribution problems, see Joachim Diederich et al., "Authorship Attribution with Support Vector Machines," *Applied Intelligence* 19 (2003): 109–23.

problem is best solved by finding the maximum separating margin between classes, i.e. the margin that leaves the maximum interclass distance, or maximizes the separability between classes. In order to construct such a margin in its constrained optimization, it considers many possible decision hyperplanes (or decision boundaries), but ultimately makes use of support vectors, which are represented in a (usually small) subset of the data items, more specifically those points of the different classes that lie closest together. These support vector points lie closest to the decision boundary, and introduce what we may call here the transitional area leading up towards it on both sides. Another aspect that has made SVM particularly powerful is its use of kernels. The intuition is that the feature vectors originally representing our training examples are augmented or replaced by kernel functions, which are a type of similarity measures (cfr. subsection A.3.5 on p. 333) between a given, current example and all other learned points. This implicit construction of additional kernel-induced features, in which training examples are not treated in isolation but are learned in the form of inner products between pairs of examples (dot products),[61] makes SVM a forerunner of more complex machine-learning methods such as neural networks.[62] The original features are, in other words, further reinforced by similarity functions which compute the similarity between the given data point and neighbouring landmarks. In that sense a kernel implicitly constructs new features, casting the existent data in a space with a higher dimensionality. This transformation of the data (by increasing dimensionality) can considerably improve the separation of classes.

### Train, Development and Test Sets

At this point, we have learned quite a few things about supervised classification. By having surveyed a handful of different classification methods, such as Naive Bayes, Burrows's Delta, *k* Nearest Neighbours and SVM, we have gained an impression of how classifiers generally behave (both from probabilistic and geometric intuitions), and in what aspects they can show their respective disparities. This has also yielded the realization that, in fact, many classifiers can be trained, and one can expect to find somewhat different results dependent on the task at hand. In this subsection, we will discuss the training process of classification, so that in the next subsection we can discuss how to evaluate the results that it yields (cfr. subsection A.3.5 on evaluation metrics).

We have already mentioned such terms as 'train' and 'test' sets. In the examples above, these respective sets can be thought of as respectively texts of known authorship (our training texts) vs. texts of unknown authorship (our test texts). In other words,

[61] Bernhard Schölkopf and Alexander J. Smola, *Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond*, Adaptive Computation and Machine Learning (Cambridge, MA: The Massachusetts Institute of Technology Press, 2002), 25.

[62] Nello Cristianini and John Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge: Cambridge University Press, 2000), 26.

—in the hypothetical situations sketched above in such figures as figure A.7— we had a test text $\gamma$, whose authorship (i.e. class, label, etc.) was sought amongst authors Suger of Saint-Denis, Peter Abelard and Bernard of Clairvaux. We assume that a classifier will be able to learn from the training samples (segments of texts of known authorship) how to recognize the behaviour of unseen samples (test samples, i.e. segments of texts of unknown authorship). In this regard, remember that for the *i* (number of) training texts, the input vectors are $< \vec{X}_1, ...\vec{X}_i >$, mapped to output labels $< Y_1, ..., Y_i >$ (see also fig. 3.4). The idea is that the model will generalize the statistical trends of input vectors $< \vec{X}_1, ...\vec{X}_i >$ for labels $< Y_1, ..., Y_i >$. Before we set out to discuss this further, remember that we have discussed sampling and segmentation of the text material in section A.3.3, and have discussed the preliminary intuitions of the machine-learning concepts mentioned here more briefly on p. 323.

### Overfitting and underfitting

Training a classifier is akin to teaching, and models can be thought of as students. This does not necessarily mean only that different students are to be approached differently in order to perform, but also that a good teacher should be able to tell a student *how* (s)he is expected to prepare and *how* to store and apply learned knowledge to be prepared for future tests. For instance, when a student is studying for an exam, a teacher might provide the student with some model questions and model answers (some training data). The wise student should not only study these individual model questions by heart, but will try to understand the theory and reasoning involved in solving those questions, so that in the future the knowledge can be applied to 'unseen' instances (test data). A student who has learned every bit of detail of the book by heart and relies solely on details, might think he or she is performing very well when still at home, telling him- or herself after solving the same model question *n* times in a row, he or she is making progress. However, when the day of the exam arrives, the student realizes that all (s)he has done was learn the model questions by heart, and that (s)he is poorly prepared for unpredictabilities. In blind and unquestioning acceptance of what was offered in the book, the student has lost sight of the 'signal,' the rules of thumb that are behind the data patterns and reveal the structures on which to rely in the future.

What I have explained in anecdotal terms above in fact refers to a well-known and frequently occurring problem in training a machine-learning classifier: overfitting and underfitting.[63] In fig. A.10 an intuition of overfitting is given. Depending on how one preprocesses the data, selects the features, chooses an algorithm, etc. one might wind up with different models yielding different decision boundaries. How to choose? Well, let us assume that in this particular case, we were lucky enough to know who wrote our test samples when we were doing the experiments. When plotting these test samples

[63] Tom M. Mitchell, *Machine Learning* (New York: McGraw-Hill, 1997), 66–9.

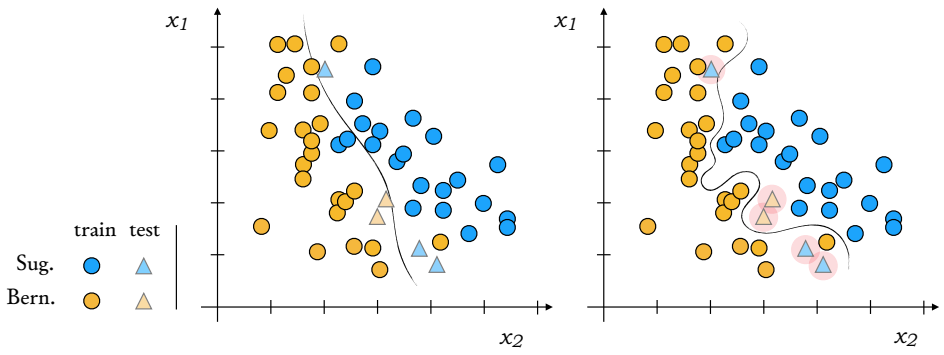



Figure A.10: Intuition of overfitting on the training data (dots). The test samples are plotted on top of the decision boundary (triangles).

over the boundary (as in fig. A.10), we can infer that the first decision boundary was in fact more apt to make future predictions than the second, and was indeed a much better fit to our data through its ability to generalize well. The test samples which we had chosen to hold out of our training set fall neatly on the correct side of the decision boundary, whereas for the right-hand side plot all test instances are falsely attributed. The left-hand side boundary ultimately proves to be a better description of the data, despite the fact that its decision boundary is slightly more relaxed than the constrained decision boundary on the right. The right-hand side line has forced itself in many angles, granting high relevance to outliers and noise, just to make certain that it would hit all the marks. This latter phenomenon is occasionally referred to as 'high bias.' 'High bias' means that, even though the training data was indeed better learned on the right-hand side and presumably obtained a very high accuracy (possibly even 100%),[64] its predictions on the test set proved that this high accuracy came at the cost of a very low flexibility and a heavily biased decision function. Ideally, a good model will find a fit that yields a balanced trade-off between what is called bias and variance.[65]

What is vital here is the idea that we can only evaluate model performance by testing its predictions against formerly labelled data. Naturally, 'in the wild,' we will not have access to such knowledge, yet at the same time it is such knowledge that we need in order to analyze our model's predictions and certify whether or not the fit is well-balanced. We cannot reasonably expect to feed all the training data to the algorithm in one turn. This enhances the risk that our algorithm basically learns the data set by heart, and it prevents us from gaining any insight whatsoever into how the model performs and why it does so as such.

[64] We will return to evaluation metrics in the next pages, see p. 342. True positives are the instances the model predicted ('positives') to be part of a class to which they indeed belong ('true').

[65] Ron Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 2 (Montreal, 1995), 1137.

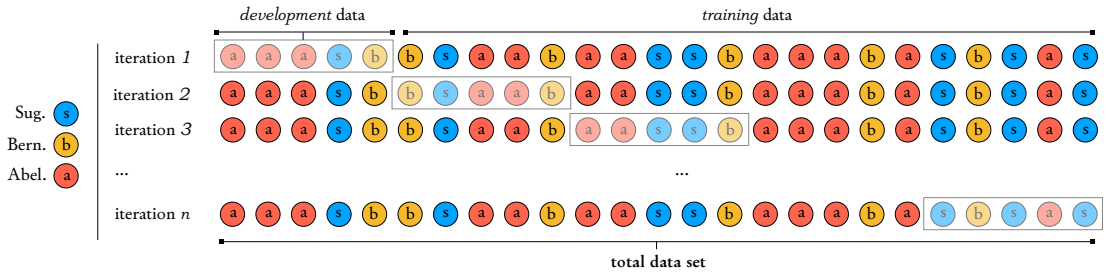



Figure A.11: Intuition of the cross-validation technique, where the majority portion of the training data is the actual training data, benchmarked for results against a held-out test set (i.e. the development set), of which we actually know the output labels.

## Cross-Validation

Overfitting is more of a symptom than an actual disease, meaning that it can be diagnosed in several ways, and often requires model-specific treatment. Some learning algorithms have specific parameters that simply need tuning in order to prevent overfitting. These fall under the header of so-called 'regularization parameters.' One can think of these as penalty scores applied to the learned weights (or coefficients). These somewhat soften the rules learned by set of parameters $\hat{\theta}$ that describe the reality of our given data best (see p. 330), which consequently yields a softer or relaxed decision boundary such as the one on the left-hand side of fig. A.10. We will return to the optimization of a classifier's parameters later on. We will here restrict ourselves to one effective and well-known procedure of preventing overfitting: cross-validation (CV).[66] CV is in fact a very simple trick that enables us to fulfill the aforementioned requirement: we need to evaluate the model by testing its predictions against preclassified data. As visualized in fig A.11, CV is precisely a simulation of 'the wild' in $k$ iterations. With this technique, we create a distribution of pairs of training and test sets out of a single data set (or population). In more transparent terms, cross-validation is a kind of sampling technique, one which was already discussed in a previous section (A.3.3, p. 319).

The total data set is divided into $k$ folds (or: batches, bins, slices, subsets, ...). In fig. A.11 such a fold consists of five text samples of an author, collected in the white bin. This number of $k$ folds equals the number of iterations by which we train our classifier. In $k$ iterations we withdraw a slice which serves as our test material, and a (commonly much larger) slice which will serve as the model's training material. This idea is commonly called a holdout procedure. It is a method by which we avoid the limitations of our dataset, which can be too small in size and/or unrepresentative as a whole to make future predictions. Another important aspect is of course that our

[66] See the important paper of Kohavi, "Cross-Validation and Bootstrap"; and for a concise state of the art on cross-validation techniques, see Ian H. Witten et al., *Data Mining. Practical Machine Learning Tools and Techniques*, 4th ed., The Morgan Kaufmann Series in Data Management Systems (Cambridge, MA: Morgan Kaufmann Publishers, 2017 (2000)), 167ff.

training data set is put to maximum use: no data gets lost as we insert it in the test set. One maximally exploits all the portions within the dataset both for training as testing.

CV simulates 'unpredictable' or 'unseen' observations in each of the $k$ training rounds, until every instance has been used exactly once for testing, allowing us to obtain a probability of how well our model will be able to predict future instances. These held-out test sets sustaining the simulation (since they are actually part of the training data, but in turn serve as test sets for evaluating the model's performance) are in the literature often referred to as 'development sets' or 'validation sets,' in order to avoid confusion with actual, unseen test sets of which the labels are truly unknown (in the examples above, such as in fig. A.9, the actual test material is of course test text $\gamma$).[67] A development set can be thought of as a stand-in for what we do not yet know for the future. The probability of a model's success in classifying each of $k$ train-test-splits is eventually expressed by evaluation metrics (to which we will return on p. 342). Ultimately, each of the $k$ cross-validated rounds are combined into the model's parameters, and the $k$ evaluation scores are averaged to a realistic estimate of how well the model has done on learning the total dataset. The maintained consensus in the state of the art is that for most problems 10 folds (i.e. $k$=10) should generally be sufficient.[68]

Cross-validation has become an indispensable part of the learning process, and in the meantime there have been introduced a number of different strategies as to how to handle the splits.[69] Two additional extensions deserve mentioning here. The first is 'stratified' CV. If one revisits fig. A.11 above, one can see that —dependent on which training partition is learned— there can be a certain bias towards the number of learned instances per class. In our hypothetical corpus which we have been entertaining for a few sections now (Abelard 13 texts; Bernard 6 texts; Suger 8 texts), we have a 48%–22%–30% proportionality. "Stratified CV" ensures that each of the folds imitates the original dataset's structure and proportionality, in order to avoid that an unrealistic bias in distribution of instances is retained by the model and represented in the prediction scores.[70] The second extension worth mentioning is the "Leave-One-Out" strategy. This particular approach departs from a number of folds that is exactly equal to the number of instances. In our particular case, this would mean that 26 texts of the 27 in our corpus are used as the training data, and consequently evaluated on 1 held-out test text (e.g. test text $\gamma$). This process is repeated until all instances have served as test data exactly once. Leave-One-Out is computationally expensive, and

[67] There are machine-learning practitioners who believe that it ever remains good practice to have one additional development set held out of the cross-validation circuit. This development set remains held out until the very last phase in the learning process to estimate the model's performance.

[68] Witten et al., *Data Mining*, 168.

[69] Sylvain Arlot and Alain Celisse, "A Survey of Cross-Validation Procedures for Model Selection," *Statistics Surveys* 4 (2010): 40–79.

[70] Ibid., 69.

| | | Author | | |
|---|---|---|---|---|
| | | Sug. | Bern. | Abel. |
| Prediction | Sug. | 5 | 2 | 0 |
| | Bern. | 3 | 3 | 2 |
| | Abel. | 0 | 1 | 11 |

(a) **True positives** ($TP$): 5 texts of Suger's writing were predicted to have been written by Suger (correct).

| | | Author | | |
|---|---|---|---|---|
| | | Sug. | Bern. | Abel. |
| Prediction | Sug. | 5 | 2 | 0 |
| | Bern. | 3 | 3 | 2 |
| | Abel. | 0 | 1 | 11 |

(b) **False negatives** ($FN$): 3 texts of Suger's writing were predicted to have been written by other authors (incorrect).

| | | Author | | |
|---|---|---|---|---|
| | | Sug. | Bern. | Abel. |
| Prediction | Sug. | 5 | 2 | 0 |
| | Bern. | 3 | 3 | 2 |
| | Abel. | 0 | 1 | 11 |

(c) **False positives** ($FP$): 2 texts by other authors were predicted to be of Suger's writing (incorrect).

| | | Author | | |
|---|---|---|---|---|
| | | Sug. | Bern. | Abel. |
| Prediction | Sug. | 5 | 2 | 0 |
| | Bern. | 3 | 3 | 2 |
| | Abel. | 0 | 1 | 11 |

(d) **True negatives** ($TN$): 17 texts written by other authors were predicted not to be of Suger's writing (correct).

Table A.3.3: Illustration of a hypothetical confusion matrix, highlighting in grey the respective $TP$, $FN$, $FP$ and $TN$ for 'class' Suger of Saint-Denis.

especially desirable if the dataset is small.[71]

### Evaluation Metrics: Accuracy, Precision, Recall and F Scores

This thesis will not discuss in full the various performance measures commonly applied to evaluate a model's accuracy of prediction. However, since such statistical jargon will nevertheless reappear throughout the thesis, it is useful to pause at what terms such as 'accuracy,' 'precision,' 'recall' and 'f1 scores' mean, allowing the uninitiated reader to gain a faint idea of what exactly these numbers can teach us. The initiated already familiar with accuracy metrics can safely skip this subsection.

A key understanding in this regard is that the accuracy of a model —despite the name— does not give sufficient information about a model's performance. That is to say, it is in fact not a reliable metric for the real performance of the classifier at all. To gain an intuition of why this is the case, consider the example of a model's predictions in table A.3.3, which is —for clarity's sake— a completely fictional outcome of results. We will return to real, more complex results as we proceed.

What is depicted in table A.3.3 is four scores the exact same hypothetical confusion matrix. A confusion matrix is a description or summary of a classifier's predictions. We see that this table is two-dimensional: one dimension represents the true class

[71] Tzu-Tsung Wong, "Performance Evaluation of Classification Algorithms by *k*-fold and Leave-One-Out Cross Validation," *Pattern Recognition* 48 (2015): 2844.

of the object (the true author), whereas the other is the classifier prediction. When summing up the values for each column in our (likewise hypothetical) training set, this set —as we know— consists of 8 texts written by Suger of Saint-Denis, 6 by Bernard of Clairvaux, and 13 by Peter Abelard. The values in the confusion matrix, on the other hand, indicate per column to which author our model attributed the texts. Consequently, the table is a very transparent and honest representation of a model's performance. However, in reality such a table often becomes too great in size to remain useful. What we want to know is, in fact, how well our model performed for each and every one of our classes (i.e. authors), in order to make more general evaluations of how well it will be able to make predictions and correct attributions in the future. Therefore, we need to look at larger trends obtained in metrics such as precision and recall, which can be captured by gathering distinctive rates within the confusion matrix that go by the name of true positives ($TP$, cfr. A.3.3a), false negatives ($FN$, cfr. A.3.3b), false positives ($FP$, cfr. A.3.3c) and true negatives ($TN$, cfr. A.3.3d).

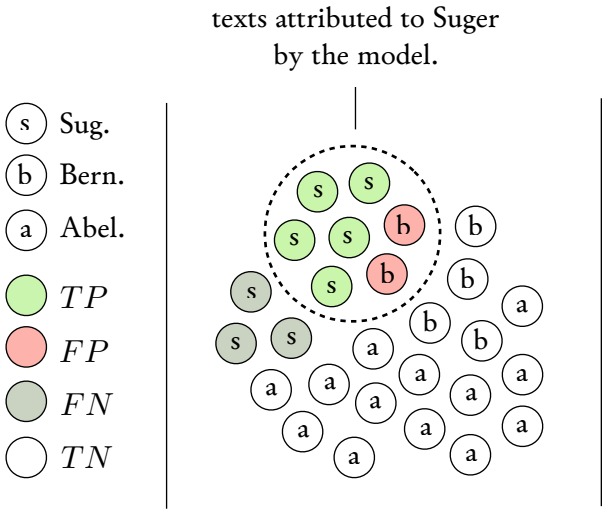



Figure A.12: Visual intuition of a model's performance, based on the confusion matrix in table A.3.3 for Suger. Each node in the figure corresponds to a text.

If one wishes to know in detail where the model is doing a rather good job, one can observe the 'true' rates, $TP$ and $TN$. These rates indicate all of the instances where the model correctly predicted 1) when a text under scrutiny was written by the actual author ($TP$) and 2) when it was not ($TN$). In the table above, which illustrates a plausible example of the performance of some model for Suger of Saint-Denis in grey highlighting, the estimator made 22 out of 27 correct decisions. 5 times it was able to correctly tell that a text was of Suger's hand (true positives), and 17 times it was able to tell that it was not (true negatives). 22 times out of 27 corresponds to an impressive percentage of 0.81%. However, one can suspect by now that evaluating a classifier solely on the basis of true rates —which corresponds to statistically determining the 'accuracy' of the model— is giving oneself an easy pass. Accuracy in essence boils

down to the following formula:

- Accuracy $(A) = \frac{TP+TN}{TP+TN+FP+FN} = \frac{5+17}{5+17+2+3} = \frac{5}{5+2}$

The 17 true negatives, included in this accuracy score, are largely informed by the fact that the model does particularly well in classifying Peter Abelard's texts, not Suger's, which —as we can moreover gather from the original set— are in fact represented in the training set with a much lower frequency (8) than texts by Abelard (13).The imbalance within the original training set has yielded a rather distorted perspective on how well this model will perform when it is confronted with a text by Suger 'in the wild.'[72]

| | Recall | Precision | Accuracy |
|---|---|---|---|
| Sug. | $\frac{5}{8}$ | $\frac{5}{7}$ | $\frac{22}{27}$ |
| Bern. | $\frac{3}{6}$ | $\frac{3}{8}$ | $\frac{19}{27}$ |
| Abel. | $\frac{11}{13}$ | $\frac{11}{12}$ | $\frac{24}{27}$ |

(a) Intuition of how recall, precision and accuracy is calculated for each individual class.

| | Recall | Precision | Accuracy |
|---|---|---|---|
| Sug. | 0.63 | 0.71 | 0.81 |
| Bern. | 0.50 | 0.38 | 0.70 |
| Abel. | 0.85 | 0.92 | 0.88 |
| **Total** | **0.66** | **0.67** | **0.80** |

(b) Recall, precision and accuracy scores averaged towards a total evaluation score for the model's predictions.

Table A.3.4: Performance measures (recall, precision and accuracy) derived from confusion matrix as demonstrated in table A.3.3.

Considering the inadequacy of accuracy for describing a model's performance, and its inadequacy of taking into account class imbalance, we need to integrate the model's weaknesses and bad calls in the evaluation (the false predictions $FP$ and $FN$). Along with this, a set of new measures is introduced: precision, recall (and F-scores). Firstly, the precision ratio in our example corresponds to expressing the following question: of all the texts which the model *did* attribute to Suger, how many were indeed written by Suger (i.e. how *precise* was the model)? This is captured through the following formula:

- Precision $(P) = \frac{TP}{TP+FP} = \frac{5}{5+2}$

Recall, on the other hand, means: of all the texts the model *should have* attributed to Suger, how many did the model recognize?

- Recall $(R) = \frac{TP}{TP+FN} = \frac{5}{5+3}$

Recall and precision are often in tension, and a good model should be doing reasonably well on both measures. An intuitive metaphor to think of the tension between precision and recall, is by thinking of a net being cast out by a shrimp fisher. If a fisher

[72] Manning, Raghavan, and Schütze, *Introduction to Information Retrieval*, 143.

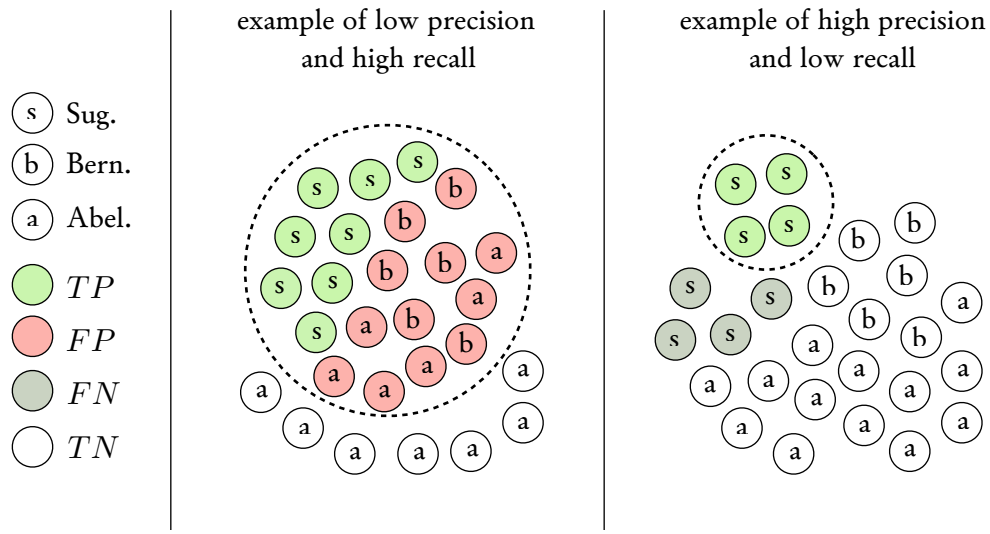



Figure A.13: Visual intuition of thresholding, i.e. what happens when the trade-off between precision and recall is imbalanced.

were to use a very small, specialized shrimp net, he would probably perform well at exclusively catching shrimp. However, the odds are high that due to the small size of the net, many shrimp will be able to elude the net. This is an example of high precision and low recall. On the other hand, a sizable fishing net will enable the shrimp fisher to catch much more shrimp, yet increases the odds that other unsought fish will get stuck in the net, making it very hard for the fisher later onwards to tell apart shrimp from other fish. This is an example of low precision and high recall.

Recall and precision only gain significance through each other: one can not know the quality of a model on the basis of the first without taking into account the latter and vice versa. An additional metric that consequently captures the combined trade-off —a type of blend of precision and recall, if you will— is the $F$ measure. The $F$ measure, or $F$ score, is the weighted harmonic (not arithmetic) mean of precision and recall.

- $F_1 = \frac{1}{\alpha\frac{1}{P}+(1-\alpha)\frac{1}{R}}$

In the metaphor of the shrimp fisher, the size of the fishing net equals what we would call the 'classification threshold.' Precision raises this threshold, whereas recall lowers it. The problem is that the threshold is not always easy to determine and can be a matter of debate, as ideally it means to provide a trade-off between the true positive rate and the false positive rate. In some cases, a high precision will be favourable to a high recall or the other way around. Therefore, it has often been proposed that experimenting with several threshold set-ups and visualizing the rates of those respective thresholds will yield better results. For instance, the ROC (Receiving Operators Characteristics) curve (found in the bottom right corner of fig. A.14) plots and evaluates many possible thresholds and calculates the number of true negatives and true

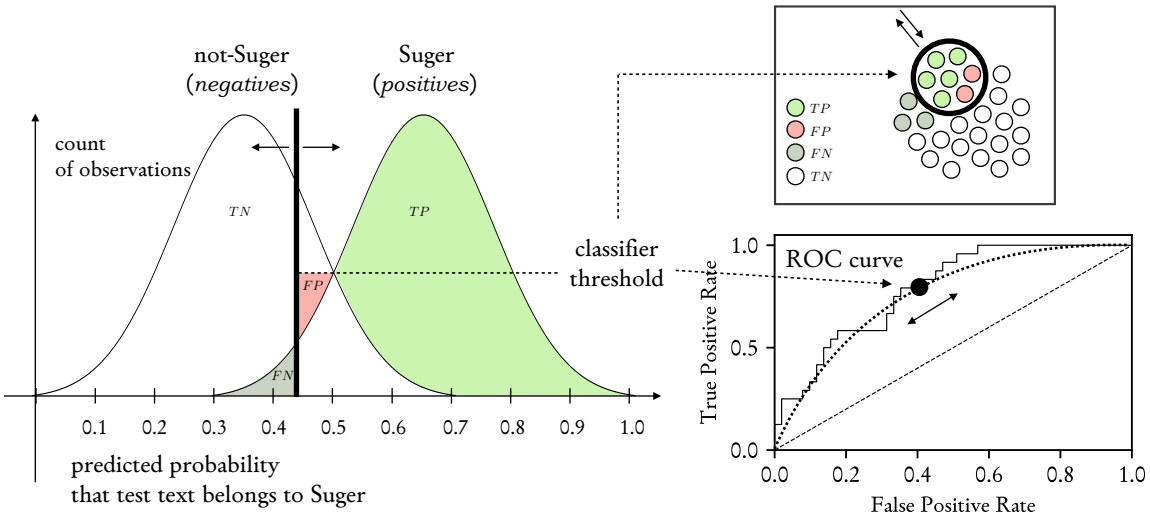



Figure A.14: Classification thresholds and ROC curve.

positives for each threshold configuration.[73] Good classifiers that can separate well between classes, will have an ROC curve that is at a good distance away from the dotted black diagonal line crossing the ROC plot. Poor estimators, however, will have an ROC curve that comes very close to the diagonal line, or even coincides with it. In essence, approximating the diagonal line means that a classifier hardly performs better than random guessing. Given that the ROC plot of our best performing estimator looks like the one in the bottom right corner of fig. A.14, an optimal trade-off between sensitivity (which is synonymous to recall) and the false positive rate is likely to be found somewhere in the middle of the curve. It is possible to quantify the ROC curve as an evaluation metric by measuring the percentage covered by the 'area under the curve,' which is often abbreviated to the AUC score. Visually speaking, the AUC score corresponds to the percentage of the box in the right bottom corner of A.14 that is coloured gray. If only half of the box were coloured, and if the ROC curve had coincided with the diagonal dotted line, this AUC score would have been 0.5, which equals random guessing. Setting the threshold, ultimately, is very much dependent on problem-specific sensibilities. The ROC curve is only a means of facilitating the interpretation of classifier performance.

Another means of visualizing different threshold settings is by generating a precision-recall curve (PR) (fig. A.15). The PR curve visualizes the gains and losses of preferring recall over precision and vice versa. Precision (y-axis) is the fraction of true positives $TP$ among the selected instances (the sum of $TP$ and $FP$). Recall (x-axis) is the fraction of true positives $TP$ among the items that should have been labelled as positive (the sum of $TP$ and $FN$). The colourbar on the bottom of fig. A.15 indicates the degree of suppleness by the algorithm, from respectively loose (low threshold, deep red), to strict (high threshold, yellow-white). A low threshold increases the odds that

[73] Witten et al., *Data Mining*, 188–9.

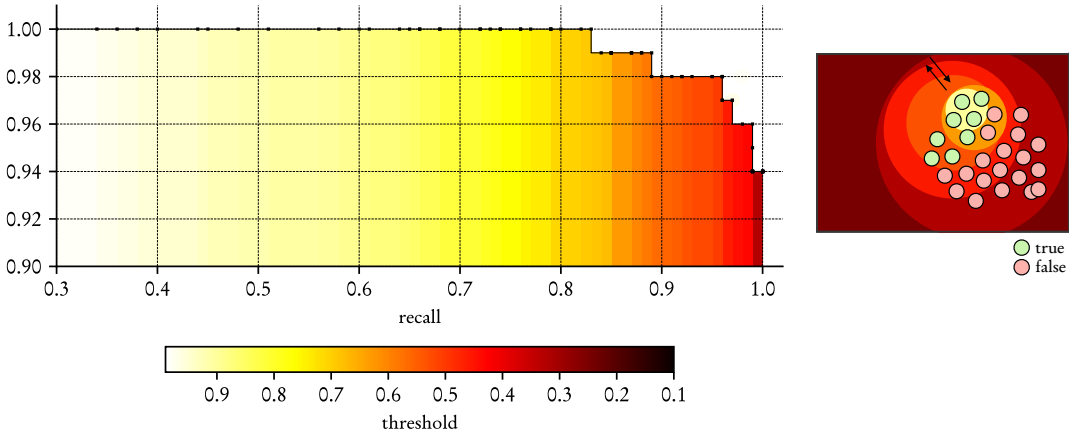



Figure A.15: PR (precision-recall) curve —or PR stairstep plot.

many instances will be falsely categorized as true. A very high threshold increases the odds that many true instances will be falsely withheld. Note that there is an important difference between the ROC and the PR curve. ROC maps out recall against the $FP$ rate, whereas a PR curve maps out recall against precision. The two are different measures and are not interchangeable.

### Parameter Optimization

Classifiers have many parameters, and changing them will change their behaviour and ultimately the evaluation metrics. It is important at this stage to stress that despite the multiplicity of methods, there are effective ways of optimizing a classifier, and of optimizing the parameters based on evaluation results. In larger set-ups, one will see that ranges of different values are applied for parametrization in order to see at which setting a certain model will perform best. One can iterate over many different settings in order to choose a model that is arguably best at the task at hand, which can avoid unreasonable extents of cherry-picking. Such iterations do not allow solely to tinker with model-specific parameters, but also to evaluate preprocessing decisions, such as sample length and number of features. Using a classifier as an additional back-up check of whether or not a proposed experimental design works for a specific authorship attribution task, is good practice in general.

## A.3.6 Visualization

In what follows, we discuss two proposed visualization methods for computational stylistics: Principal Components Analysis (PCA) and Networks. We likewise address the main issue referred to in the main text on p. 96: visualization in data analysis easily distracts from the underlying data instead of describing it. It can —deliberately or not— prove subjective and misleading.

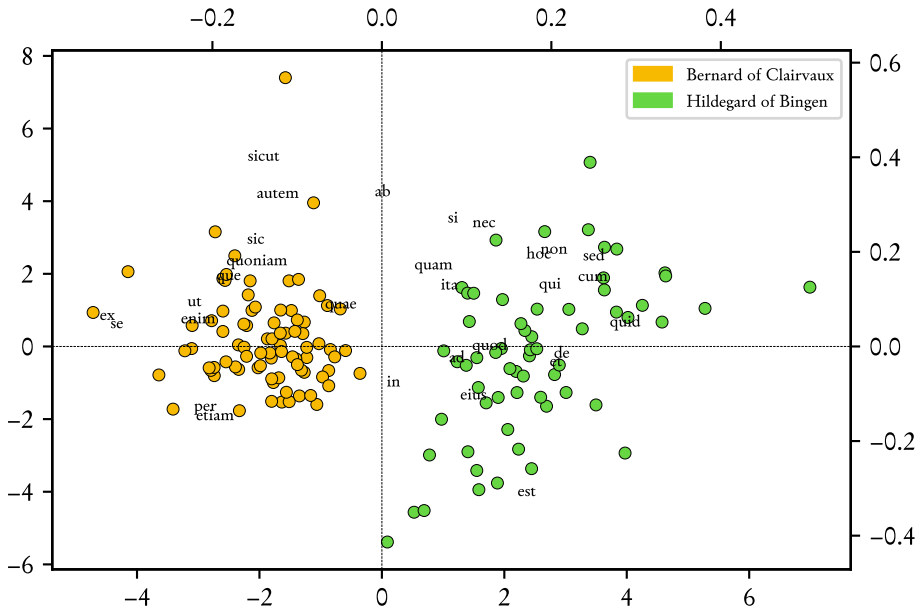



Figure A.16: PCA cluster plot for Bernard of Clairvaux's *Sermones Super Cantica Canticorum* and Hildegard of Bingen's *Liber Divinorum Operum*. Rehearsal of fig. 2.3 (p. 68. Explained variance: 29.72%. Number of words: 30. Sample size : 2,000 *w*).

## Principal Components Analysis (PCA)

PCA is an unsupervised technique that allows to reduce a multivariate or multidimensional data set of many features, such as function word frequencies, to merely two or three principal components (PC's), which disregard inconsequential information (or noise) in the data set and reveal its important dynamics.[74] The assumption is that the main PC's, the axes in the plot, point in the direction of the most significant change in the data, so that clustering and outliers become clearly visible. Each word in the feature vector is assigned a weighting, or loading, which reflects whether or not a word correlates highly with a PC and therefore gains importance as a discriminator in writing style. In a plot (see fig. A.16), the loadings or features that overlap with the clustered texts of a particular author are the preferred stylistic features of that author. PCA is built to find the most meaningful variance of observations along the axes of its principal components. In this sense it is not always interested in finding links between candidates, as *k*-NN is for example (A.3.5), but rather in finding links between variables (i.e. features). A disadvantage to PCA is that it can never explain all the variance of the data, since it purposefully disregards many features and dimensions that it finds insignificant. It also has the tendency to produce somewhat nebulous scatter plots when texts are stylistically entangled, or if the data set is large in general.

[74] For an elaborate explanation of PCA and its applicability to stylometry, see José Nilo G. Binongo and M. Wilfrid A. Smith, "The Application of Principal Components Analysis to Stylometry," *Literary and Linguistic Computing* 14, no. 4 (1999): 446–66.

### Network Graph Analysis

In its simplest form, "A network is [...] a collection of points joined together in pairs by lines. In the jargon of the field the points are referred to as vertices or nodes and the lines are referred to as edges."[75] "Nodes" correspond to data points, and they are mapped in a space through force-directed graph drawing on the basis of connections which are called edges.[76] This can be thought of as a game of attraction and repulsion in some kind of magnetic force field, where similar elements are drawn to each other, and dissimilar elements pushed away. The thicker the connection and the closer the nodes are mapped in a space, the more these nodes resemble each other in terms of the predefined, quantified relationship. Network theory is in fact a broad field that has wide applicabilities in computer science, biology, social sciences, mathematics, physics, and many other areas. It is only recently that its usefulness for computational stylistics has taken flight, by such computational stylisticians as Moretti,[77] Jockers,[78] Rybicki,[79] and Eder.[80] What is interesting about networks, is that the connections between different data points can be defined and stored in a variety of ways. This makes networks very flexible visualization methods, but the concern has also been raised whether networks have not been applied in situations where they are unsuited.[81] Since edges have the ability to represent many different types of algorithmic relationships at once, the idea of so-called consensus networks have attracted a fair amount of attention lately.[82] The main idea is that a consensus is established in the edges' links along a series of different algorithms, parameters and feature frequency strata. For a large extent this eliminates the possibility that the analyst weights relationships between nodes him- or herself, or cherry-picks results that look most convenient. Networks' ability to combine much information in edges has also made it, in opposition to PCA, for instance, suitable for visualizing large-scale stylistic connections.

In its most typical guise, networks are simply a variant to the unsupervised $k$ nearest neighbours algorithm (connected to the supervised $k$ nearest neighbours classifier in A.3.5 (p. 335)). The idea is that for each document vector $\vec{X}$, the $k$ nearest neighbours

[75] Mark E. J. Newman, *Networks. An Introduction* (Oxford, UK: Oxford University Press, 2010), 1.

[76] Mathieu Jacomy et al., "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software," *Plos One* 9, no. 6 (2014): 1–12.

[77] Moretti applied network analysis to study the plot of *Hamlet*, in Franco Moretti, "Network Theory, Plot Analysis," *New Left Review* 68 (2011): 80–102.

[78] Matthew Lee Jockers, *Macroanalysis: Digital Methods and Literary History*, Topics in the Digital Humanities (Champaign, IL: University of Illinois Press, 2013), 163ff.

[79] Jan Rybicki, Maciej Eder, and David L. Hoover, "Computational Stylistics and Text Analysis," in *Doing Digital Humanities*, ed. Constance Crompton, Richard J. Lane, and Ray Siemens (London and New York: Routledge, 2016), 123–44.

[80] Eder, "Visualization in Stylometry."

[81] "Networks *can* be used on any project. Networks *should* be used on far fewer," see Scott B. Weingart, "Demystifying Networks, Parts I & II," *Journal of Digital Humanities* 1, no. 1 (2011), `http://journalofdigitalhumanities.org/1-1/demystifying-networks-by-scott-weingart/`.

[82] As it is termed in Eder, "Visualization in Stylometry."

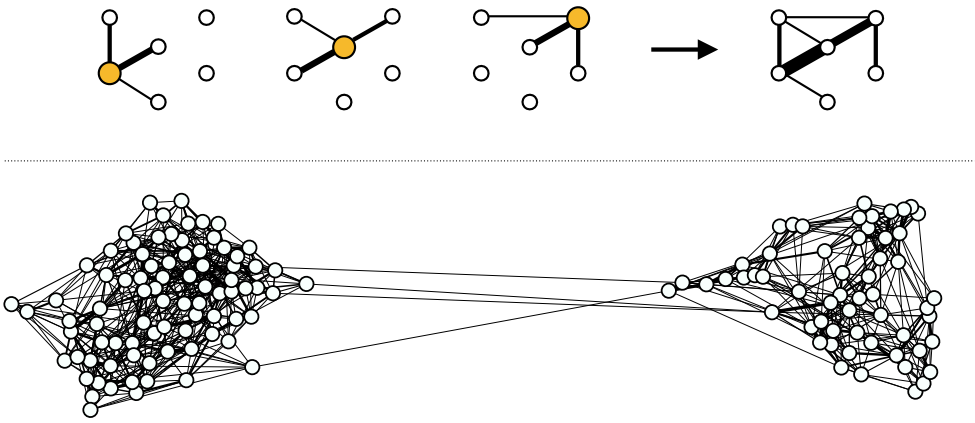


Figure A.17: Intuition of force-directed graph drawing. The top corner of the figure (inspired to a great extent by Eder, "Visualization in Stylometry," 58) shows how for each node the *k* closest neighbours are chosen. The connections between the nodes are stored within edge weights, which consequently adjust the thickness of the edge. The bottom corner of the figure shows a more complex network graph with 144 nodes and 788 edges.

are calculated according to a defined distance metric (e.g. euclidean, manhattan, cosine, ... see A.3.5 (p. 333)). Consequently, the measured distance can serve as a normalized weight. The smaller the measured distance, the higher the weight thickening the edge's line. This can be achieved, for instance, by normalizing the distances to a (1,0) range. The interesting aspect of networks, however, is that one can search across different feature frequencies and distance parameters, and consequently average the result within a defined edge. This is Eder's proposed consensus technique.[83]

It should be noted that networks always find relationships, as it is very much a closed game. It is designed to link candidates to one another in terms of distance (every text sample needs to find its respective *k* neighbours) and can presuppose ties that are rather coincidental or nonexistent (for example, in the case of outliers). Therefore, a network visualization can be biased because of a misleading directionality. Another disadvantage that should be raised here is that force-directed graphs are meant to be, above all, "aesthetically pleasing with more emphasis on symmetries and non-overlapping nodes."[84] This does not mean that the visualization therefore becomes useless, but one should analyze them with the awareness that the underlying force-directed algorithms which are responsible for mapping out the nodes in a space, might contort the original proportions of the data for the sake of overview or transparency. Not precision and accuracy are the main goals of these visualizations, but intuition and a familiarity with the data.

[83] Eder, "Visualization in Stylometry."

[84] Devangana Kokhar, *Gephi Cookbook* (Birmingham, UK: Packt Publishing, 2015), 65.

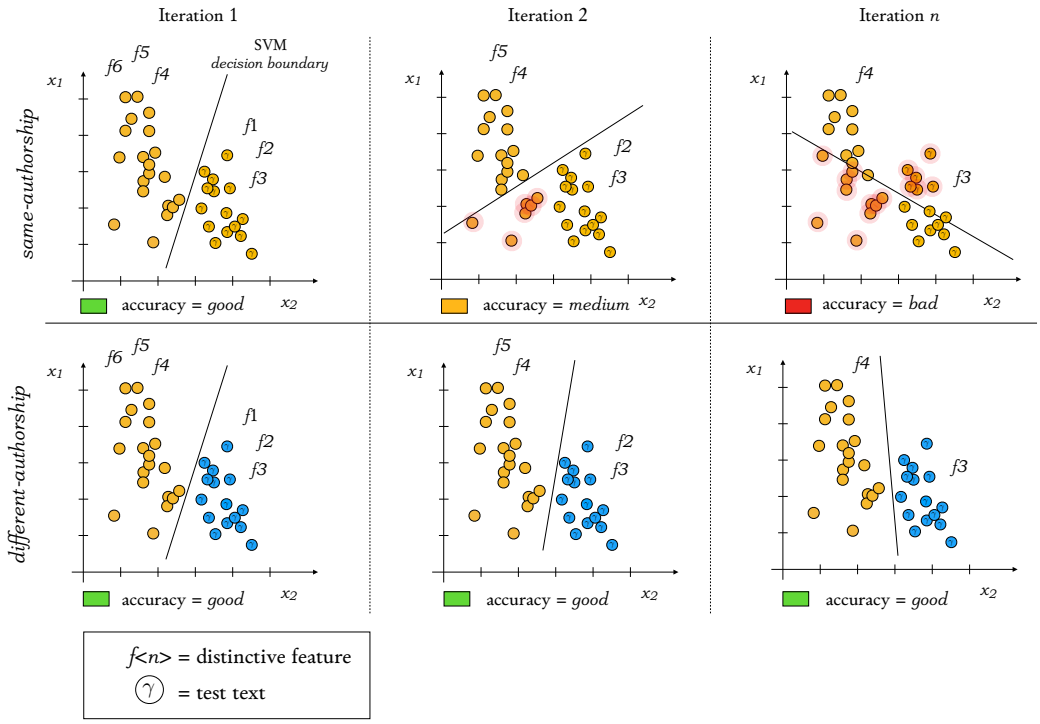



Figure A.18: Intuition of the *unmasking* method. The top row is an example of how the accuracy of an SVM trained on texts of similar authorship will fail dramatically after removing the most distinctive features (*f*<*n*>) in *n* iterations. When removing the best segregators, the separator (decision boundary) soon fails at distinguishing the learned classes well, a failure which suggests that the classes are in fact inseparable because they are written by the same author. The bottom row, on the other hand, shows how the vector differences between two works of different authorship are not superficial, yet structurally consistent down to lower frequency strata, wherefore the accuracy decreases much less quickly.

## A.3.7 *Unmasking*

*Unmasking* was a method first proposed by Koppel and Schler in 2004.[85] In many regards, one can think of *unmasking* as a kind of perverse classification experiment, exploring the limits of machine-driven aptitude for segregation. Instead of receiving classification error as failure, *unmasking* turns a disadvantage to its advantage: if a classifier is shown to be unsuccessful in segregating two documents, then that classifier can consequently be proven successful in recognizing their identicality.

In fig. A.18, one can see quite well that stylistic variability can be deceiving in a visualization, where the concept of 'distance' is purposefully made to look ambiguous: in the scenario depicted in the top row, a classifier separates works of same authorship; in the bottom row, a classifier separates works of different authorship. Indeed, there is a calculable distance for both scenarios, which —if so desired— can be separated by a classifier. Yet we do not have an estimation of reliability concerning the classification

[85] Moshe Koppel and Jonathan Schler, "Authorship Verification as a One-Class Classification Problem," in *Proceedings of the 21st International Conference on Machine Learning* (New York: ACM Press, July 2004), s.p.

made: is this classifier distinguishing these works because of different authorship? Not because two documents are separable, do they necessarily indicate different authorship. This is a much encountered problem in so-called 'closed games,' where we presuppose that a text was written by some author that we have available in our data set. The success or failure of such classifications hinges on the presupposition that the author of some anonymous test text $\gamma$ is included in our line-up at all. In real-world examples, we can rarely assume that this will be the case (with most acute examples in forensic applications of style detection, e.g. bomber or suicide notes). Authorship verification, then, comes with a realization that we need to build in "an additional category label: 'None of the above.'"[86] We need some kind of benchmark which, when crossed, should inform us on the classification's (un)reliability. Or, as Kestemont et al. phrased it in their test runs of *unmasking* across genres: "The unmasking approach does not test whether a stylistic model can be built, distinguishing between two texts, since this is often all too easy. Rather, it tests the robustness of this model by deliberately impairing it over a number of iterations, each time removing those features that are most discriminative between the two texts."[87] In other words: we need to learn more about the learning of classifiers, meta-learning.

The *unmasking* method proposes a kind of bootstrapping in which the best selected features to distinguish between two sets of documents are iteratively dropped. The design of the *unmasking* method is the following. A linear SVM classifier is trained (10-fold cross validation) to distinguish between test texts $\gamma$ and a set of texts by some author, who is either the true author (i.e. the top row of fig. A.18) or the false author (i.e. the bottom row of fig. A.18). Koppel and Schler propose feature vectors of some 250 words. In the classification training process, a feature selection algorithm is called up to rank the features. In each iterative step, the 3 most strongly-weighted positive features and the 3 most strongly-weighted negative features are eliminated. The result is the well-known *unmasking* degradation curve (A.19), where the main intuition is the following: "if [set 1] and [set 2] are by the same author, then whatever differences there are between them will be reflected in only a relatively small number of features, despite possible differences in theme, genre and the like."[88] Removing this relatively small number of features that are distinctive in iterative steps, conjures up bad decision boundaries, and a quick and dramatic drop of accuracy, which can be represented in a degrading curve. If this dramatic drop occurs, we are commonly handling texts of same authorship. The question rises, then again, how can we be absolutely sure of this?

Koppel and Schler's idea is to collect many degradation curves (which can easily

[86] Justin A. Stover and Mike Kestemont, "The Authorship of the *Historia Augusta*: Two New Computational Studies," *Bulletin of the Institute of Classical Studies of the University of London*, 2016, 145.

[87] Mike Kestemont et al., "Cross-Genre Authorship Verification Using Unmasking," *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 342.

[88] Koppel and Schler, "Authorship Verification as a One-Class Classification Problem."

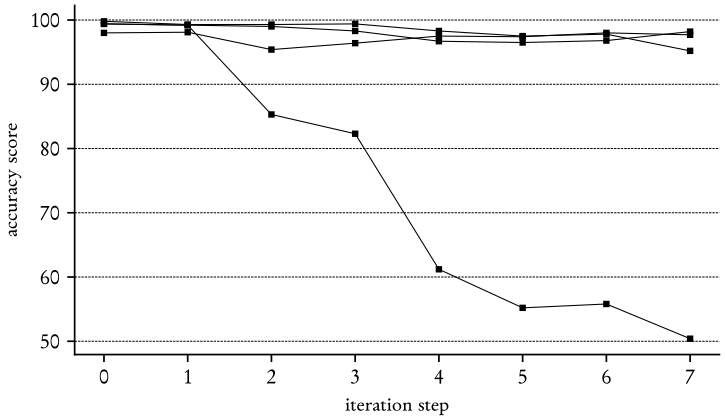



Figure A.19: Example of a hypothetical *unmasking* degradation curve, in which classification accuracy is plotted as a function of the iteration index. For each iteration step the 3 most strongly-weighted positive features and the 3 most strongly-weighted negative features are eliminated. The 10-fold cross-validation accuracy score decreases significantly when the classifier was trained on texts of same authorship (the single obvious degradation curve sinking to ±55%), as opposed to classifiers trained on texts of different authorship (the 3 top curves, which barely budge).

be generated by using labelled training data), and then use such degradation vectors as training data with labels <*same-author*> or <*different-author*> for a new classifier (meta-learning or 'ensemble learning').[89] The degradation curve of some unknown author can subsequently be fed to this algorithm. This learned classifier could then in its own turn be applied in cases of unknown authorship in the wild, in order to distinguish between classes either <*same-author*> or <*different-author*>.

[89] "It would be equally productive if new methods like unmasking [...] and variants would be framed as instances of stacked classiers and ensemble learning, which they are, thereby providing more clarity," see Walter Daelemans, "Explanation in Computational Stylometry," in *Computational Linguistics and Intelligent Text Processing*, ed. Alexander Gelbukh, vol. 7817 (Berlin, 2013), 451–462.

# A.4 Addenda to Chapter 4

| Title | Location in *PL* 144 |
|---|---|
| *In nativitate S. Ioannis Baptistae* | 627 |
| *In natali apostolorum Petri et Pauli* | 649 |
| *In natali S. Benedicti de evangelio* | 548 |
| *In festivitate S. Mariae Magdalenae* | 660 |
| *In festivitate S. Petri ad vincula* | 646 |
| *In assumptione B. Mariae* | 717 |
| *In nativitate B. Mariae* | 736 |
| *In exaltatione S. crucis* | 761 |
| *In festivitate angelorum* | 794 |
| *In dedicatione ecclesiae* | 897 |
| *In festivate S. Victoris* | 732 |
| *In festivitate omnium sanctorum* | 811 |
| *In festivitate S. Martini* | 815 |
| *In festivitate S. Andreae* | 828 |
| *In festivitate B. Nicholai* | 835 |
| *In festivitate B. Mariae* | 557 |
| *In vigilia nativitatis* | 839 |
| *In nativitate Domini* | 847 |
| *In festivitate B. Stephani* | 853 |

Table A.4.1: Sermons by Nicholas of Montiéramey, collected among those of Peter Damian in *PL* 144.

| 1-25 | 26-50 | 51-75 | 76-100 | 101-125 | 126-150 |
|---|---|---|---|---|---|
| et | que | sine | tamquam | iuxta | donec |
| in | sibi | nam | ante | verum | cito |
| qui | pro | ita | utique | itaque | nimirum |
| non | enim | magis | contra | pote | numquam |
| is | vel | vero | nullus | secundum | cur |
| hic | ex | apud | igitur | multum | plane |
| quod | autem | bene | certe | quando | absque |
| ego | ne | tantus | aliquis | alter | quatenus |
| sed | per | inter | dum | ibi | proinde |
| de | aut | immo | semper | tunc | ceterum |
| ut | tamen | propter | videlicet | sane | longe |
| ad | iam | quippe | quidam | uterque | pariter |
| ille | quo | quoque | quisquis | nemo | facile |
| si | quidem | idem | siquidem | omnino | at |
| ab | sic | ac | sub | sive | inde |
| cum | iste | solum | satis | profecto | simul |
| quis | alius | denique | usque | nonne | ubique |
| suus | nisi | talis | quantum | prius | tandem |
| ipse | super | quoniam | numquid | porro | ideo |
| quam | etiam | adhuc | neque | ample | coram |
| meus | tam | atque | an | alioqui | huiusmodi |
| quia | ergo | quantus | post | vere | iterum |
| nec | ubi | quomodo | unquam | utinam | rursus |
| nos | sicut | etsi | quasi | libenter | quisque |
| noster | nunc | unde | minus | interim | parve |

Table A.4.2: Most frequent function words for figs. 4.1–4.2 (the letters).

| 1-25 | 26-50 | 51-75 | 76-100 | 101-125 | 126-150 |
|---|---|---|---|---|---|
| et | nos | nam | uterque | iuxta | seipse |
| in | per | quoniam | aliquis | quisquis | item |
| qui | ex | inter | tunc | videlicet | quicumque |
| non | autem | denique | solum | apud | an |
| hic | noster | magis | sane | profecto | donec |
| is | que | nunc | quando | scilicet | certe |
| sed | vel | unde | igitur | prius | vere |
| ad | ergo | quidam | ante | nemo | quisque |
| ille | quidem | sine | talis | parve | absque |
| quod | tamen | propter | post | porro | interim |
| ut | iste | quasi | bene | plane | unquam |
| de | pro | tam | nullus | ibi | numquam |
| ego | iam | atque | sub | contra | quantum |
| cum | alius | quomodo | omnino | immo | pote |
| suus | ne | quoque | usque | nonne | prorsus |
| ab | etiam | tamquam | semper | at | semetipse |
| si | aut | ac | quippe | nimirum | pariter |
| ipse | sic | tantus | sive | nihilominus | amen |
| quis | sicut | idem | alter | primum | proinde |
| quia | quo | neque | minus | propterea | satis |
| sibi | nisi | utique | etsi | verum | huiusmodi |
| meus | vero | adhuc | inde | nempe | numquid |
| enim | super | dum | siquidem | una | hinc |
| nec | ita | quantus | itaque | multum | aliquando |
| quam | ubi | secundum | ideo | longe | prae |

Table A.4.3: Most frequent function words for figs. 4.3–4.4 (the sermons).

| index | *SBO* | index | *SBO* |
|---|---|---|---|
| 1 | Epp. 1.1-13 | 9 | Epp. 89.3ff., 91, 95, 96, 98, 102, 104, 106, 107.1-3 |
| 2 | Epp. 1.13ff., 2.1-12 | 10 | Epp. 107.3ff., 111, 113.1-2 |
| 3 | Epp. 2.12ff., 8, 11.1-5 | 11 | Epp. 113.2ff., 114, 117, 118, 119, 124, 125, 126.2 |
| 4 | Epp. 11.5ff., 12, 24, 25, 42, 65, 67.1-1 | 12 | Epp. 126.2ff., 127.1-2 |
| 5 | Epp. 67.1ff., 68, 69, 70, 72.1-3 | 13 | Epp. 127.2ff., 129, 130, 131, 132, 133, |
| 6 | Epp. 72.3ff., 73, 78.1-11 | 14 | Epp. 143.3ff., 150, 152, 156, 158, 159, 168, |
| 7 | Epp. 78.11ff., 79, 82, 83, 85, 87.1-2 | | 169, 178.1-5 |
| 8 | Epp. 87.2ff., 88, 89.1-3 | | |

Table A.4.4: Description of sample contents (3,000 *w*) for Bernard's *intra* corpus (*brevis* publication) in figs. 4.1–4.2.

| index | *SBO* | index | *SBO* |
|---|---|---|---|
| 1 | Epp. 3, 5, 6, 7.1-9 | 14 | Epp. 188.1ff., 189, 190, 191, 192, 193, 194, 195, 196.1 |
| 2 | Epp. 7.9ff., 9, 10 | 15 | Epp. 196.1ff., 197, 198, 99, 200, 201, 202, 203, 204, 205, 206, 207 |
| 3 | Epp. 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23.1-4 | 16 | Epp. 208, 209, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221.1 |
| 4 | Epp. 23.4ff., 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37.1 | 17 | Epp. 221.1ff., 222, 223, 224, 225, 226, 227 |
| 5 | Epp. 37.1ff., 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49 | 18 | Epp. 228, 229, 230, 231, 232, 233, 234, 235, 236.1 |
| 6 | Epp. 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 66 | 19 | Epp. 236.1ff., 237, 238, 239, 240.2 |
| 7 | Epp. 71, 74, 75, 76, 77, 80, 81, 84, 86, 90, 92, 93, 94, 97.1 | 20 | Epp. 240.2ff., 241, 242, 243, 244.3 |
| 8 | Epp. 97.1ff., 99, 100, 101, 103, 105, 108.1 | 21 | Epp. 244.3ff., 245, 246, 247, 248, 249, 250, 251, 252, 253.1 |
| 9 | Epp. 108.1ff., 109, 110, 112, 115, 116, 120, 121, 122, 123, 128 | 22 | Epp. 253.1ff., 255, 256, 257, 258, 259 |
| 10 | Epp. 134, 135, 137, 138, 140, 142, 144, 145, 146, 147.1-2 | 23 | Epp. 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275 |
| 11 | Epp. 147.2ff., 148, 149, 151, 153, 154, 155, 157, 160, 161, 162, 163, 164, 165, 166.1 | 24 | Epp. 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286 |
| 12 | Epp. 166.1ff., 167, 170, 171, 172, 173, 174, 175 | 25 | Epp. 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301.1-2 |
| 13 | Epp. 176, 177, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188.1 | | |

Table A.4.5: Description of sample contents (3,000 *w*) for Bernard's *intra* corpus (*perfectum* publication) in figs. 4.1–4.2.

| index | *SBO* | index | *SBO* |
|---|---|---|---|
| 1 | Epp. 311, 312, 313, 314, 315, 316, 317, 318, 319, 322.1 | 3 | Epp. 397.4ff., 398, 399, 400, 406, 407, 410, 411, 413, 431, 432, 441, 449 |
| 2 | Epp. 322.1ff., 324, 359, 391, 392, 394, 396, 397.1-4 | | |

Table A.4.6: Description of sample contents (3,000 *w*) for Bernard's *extra* corpus (pre-1140) in figs. 4.1–4.2.

| index | *SBO* | index | *SBO* |
|---|---|---|---|
| 1 | Epp. 320, 321, 323, 327, 330, 331, 332, 333, 334, 335, 336, 338.1 | 3 | Epp. 356, 357, 358, 360, 362, 385, 393, 416 |
| 2 | Epp. 338.1ff., 339, 340, 341, 342, 346, 347, 348, 349, 350, 351, 353, 354, 355 | 4 | Epp. 433, 434, 435, 436, 437, 438, 439, 440, 447, 505, 520, 523, 525 |

Table A.4.7: Description of sample contents (3,000 *w*) for Bernard's *extra* corpus (1140–45) in figs. 4.1–4.2.

| index | *SBO* | index | *SBO* |
|---|---|---|---|
| 1 | Epp. 328, 329, 345, 361, 363, 364, 365.1-2 | 3 | Epp. 377.2ff., 378, 379, 380, 381, 382, 383, 384, 387, 389, 390, 401, 402, 403.1-2 |
| 2 | Epp. 365.2ff., 366, 367, 368, 369, 370, 371, 372, 374, 375, 376, 377.1-2 | 4 | Epp. 403.2ff., 409, 417, 418, 419, 420, 421, 451, 455, 457, 458, 508, 509, 515, 521 |

Table A.4.8: Description of sample contents (3,000 *w*) for Bernard's *extra* corpus (post-1145) in figs. 4.1–4.2.

| index | *PL* (vol.:col.) | index | *PL* (vol.:col.) |
|---|---|---|---|
| 1 | Ep. 1 (196: 1593a-1594b); Ep. 2 (196: 1594b-1596a); Ep. 3 (196: 1596b-1597b); Ep. 4 (196: 1597b-1598c); Ep. 5 (196: 1598d-1600a); Ep. 6 (196: 1600b-1601b); Ep. 7 (196: 1601c-1601d) | 5 | Ep. 35 (196: 1626d-1631a); Ep. 36 (196: 1631b-1632c); Ep. 38 (196: 1632c-1635b) |
| 2 | Ep. 7 (196: 1601d-1603a); Ep. 8 (196: 1603b-1605a); Ep. 9 (196: 1605b-1605d); Ep. 10 (196: 1606a-1607d); Ep. 11 (196: 1608a-1608c); Ep. 12 (196: 1608c-1609a); Ep. 15 (196: 1609b-1610a) | 6 | Ep. 38 (196: 1635b- 1636c); Ep. 40 (196: 1636d-1639d); Ep. 41 (196: 1640a-1640b); Ep. 42 (196: 1640c-1641c); Ep. 43 (196: 1641c-1643b) |
| 3 | Ep. 15 (196: 1610a-1610c); Ep. 16 (196: 1610d-1613c); Ep. 17 (196: 1613d-1616a); Ep. 18 (196: 1616b-1617c); Ep. 19 (196: 1617d-1618a) | 7 | Ep. 43 (196: 1643b-1644a); Ep. 44 (196: 1644a-1645a); Ep. 45 (196: 1645b-1646d); Ep. 46 (196: 1647a-1648c); Ep. 47 (196: 1648d-1649a); Ep. 50(196: 1649c-1650c); Ep. 51 (196: 1651a-1651d) |
| 4 | Ep. 23 (196: 1618c-1619a); Ep. 27 (196: 1619c-1620a); Ep. 29 (196: 1620b-1621c); Ep. 31 (196: 1621d-1622d); Ep. 32 (196: 1623a-1623c); Ep. 33 (196: 1623c-1625c); Ep. 34 (196: 1625d-1626c) | | |
| 1 | Sm. 69 (144: 897c-902b); Sm. 43 (144: 732b-735b) | 8 | Sm. 23 (144: 629c-637a); Sm. 27 (144: 649a-649b) |
| 2 | Sm. 43 (144: 735c-736b); Sm. 55 (144: 811c-815c); Sm. 56 (144: 815d-818b) | 9 | Sm. 27 (144: 649c-652c); 9.42. hom. (144: 548c-553a) |
| 3 | Sm. 56 (144: 818c-822d); Sm. 58 (144: 828d-832a) | 10 | 9.42. hom. (144: 553b); Sm. 29 (144: 660b-666a); Sm. 26 (144: 646b-647d) |
| 4 | Sm. 58 (144: 832b-834c); Sm. 59 (144: 834d-838d); Sm. 11 (144: 557a-558a) | 11 | Sm. 26 (144: 648a-649a); Sm. 40 (144: 717a-722c); Sm. 44 (144: 736b-737b) |
| 5 | Sm. 11 (144: 558b-563a); Sm. 60 (144: 839b-841d) | 12 | Sm. 44 (144: 737c-740d); Sm. 47 (144: 761c-765c) |
| 6 | Sm. 60 (144: 842a-846a); Sm. anonym. (144: 848b-851d) | | |
| 7 | Sm. anonym. (144: 852a-853b); Sm. 62 (144: 853b-857c); Sm. 23 (144: 627b-629b) | | |

Table A.4.9: Description of sample contents (3,000 *w*) for Nicholas's letters and sermons in figs. 4.1–4.2.

| index | *SBO* | index | *SBO* |
|---|---|---|---|
| 1 | Sm. 1.1-7 | 25 | Sm. 33.4ff., 34.1–3 |
| 2 | Sm. 1.7ff., 2.1-6 | 26 | Sm. 34.3ff., 45.1–5 |
| 3 | Sm. 2.6ff., 3.1-4 | 27 | Sm. 45.5ff., 47, 48, 49, 50.1–3 |
| 4 | Sm. 3.4ff., 4.1-2 | 28 | Sm. 50.3ff., 51, 52, 53, 54 |
| 5 | Sm. 4.2ff., 5.1-4 | 29 | Sm. 55, 56, 57.1 |
| 6 | Sm. 5.4ff., 8.1 | 30 | Sm. 57.1ff., 58, 59, 60, 61.1 |
| 7 | Sm. 8.1-8 | 31 | Sm. 61.1ff., 63, 64, 65, 66, 67 |
| 8 | Sm. 8.8ff., 10.1-2 | 32 | Sm. 69, 70, 71, 72.1–3 |
| 9 | Sm. 10.2ff., 11, 12.1-3 | 33 | Sm. 72.3ff., 73, 74, 75, 76, 77 |
| 10 | Sm. 12.3ff., 13, 14.1-4 | 34 | Sm. 78, 79, 80, 81, 82, 84, 85 |
| 11 | Sm. 14.4ff., 15.1-4 | 35 | Sm. 86, 87, 88.1 |
| 12 | Sm. 15.4ff., 16.1-6 | 36 | Sm. 88.1ff., 89, 90.1–5 |
| 13 | Sm. 16.6ff., 17.1-6 | 37 | Sm. 90.5ff., 91.1–7 |
| 14 | Sm. 17.6ff., 18, 19.1 | 38 | Sm. 91.7ff., 92, 93, 94.1 |
| 15 | Sm. 19.1ff., 20.1-3 | 39 | Sm. 94.1ff., 95, 96.1–3 |
| 16 | Sm. 20.3ff., 22.1-6 | 40 | Sm. 96.3ff., 97, 98 |
| 17 | Sm. 22.6ff., 23.1-4 | 41 | Sm. 99, 101, 102, 103.1–3 |
| 18 | Sm. 23.4ff., 24.1-4 | 42 | Sm. 103.3ff., 105, 106, 107.1–2 |
| 19 | Sm. 24.4ff., 25.1-8 | 43 | Sm. 107.2ff., 108, 109, 110, 111.1–4 |
| 20 | Sm. 25.8ff., 26, 27.1-2 | 44 | Sm. 111.4ff., 112, 113, 115, 116, 117, 118 |
| 21 | Sm. 27.2ff., 28.1-2 | 45 | Sm. 119, 120, 121, 122, 123, 124.1–2 |
| 22 | Sm. 28.2ff., 29.1-2 | 46 | Sm. 124.2ff., 125.1–3 |
| 23 | Sm. 29.2ff., 30, 31.1-3 | | |
| 24 | Sm. 31.3ff., 32, 33.1-4 | | |

Table A.4.10: Description of sample contents (1,500 *w*) for Bernard of Clairvaux's *Sermones de diversis* in figs. 4.3–4.4.

| index | *SBO* | index | *SBO* |
|---|---|---|---|
| 1 | Sm. 1.1–1.11 | 56 | Sm. 42.6–43.2 |
| 2 | Sm. 1.11–2.8 | 57 | Sm. 43.2–44.5 |
| 3 | Sm. 2.8–4.1 | 58 | Sm. 44.5–45.6 |
| 4 | Sm. 4.1–5.6 | 59 | Sm. 45.6–46.4 |
| 5 | Sm. 5.6–6.6 | 60 | Sm. 46.4–47.3 |
| 6 | Sm. 6.6–7.7 | 61 | Sm. 47.3–48.3 |
| 7 | Sm. 7.7–8.6 | 62 | Sm. 48.3–49.1 |
| 8 | Sm. 8.6–9.4 | 63 | Sm. 49.1–50.1 |
| 9 | Sm. 9.4–10.4 | 64 | Sm. 50.1–50.8 |
| 10 | Sm. 10.4–11.2 | 65 | Sm. 50.8–51.8 |
| 11 | Sm. 11.2–12.1 | 66 | Sm. 51.8–52.6 |
| 12 | Sm. 12.1–12.9 | 67 | Sm. 52.6–53.6 |
| 13 | Sm. 12.9–13.4 | 68 | Sm. 53.6–54.4 |
| 14 | Sm. 13.4–14.1 | 69 | Sm. 54.4–54.10 |
| 15 | Sm. 14.1–14.8 | 70 | Sm. 54.10–56.1 |
| 16 | Sm. 14.8–15.6 | 71 | Sm. 56.1–57.2 |
| 17 | Sm. 15.6–16.4 | 72 | Sm. 57.2–57.9 |
| 18 | Sm. 16.4–16.14 | 73 | Sm. 57.9–58.6 |
| 19 | Sm. 16.14–17.7 | 74 | Sm. 58.6–59.1 |
| 20 | Sm. 17.7–18.6 | 75 | Sm. 59.1–59.9 |
| 21 | Sm. 18.6–19.7 | 76 | Sm. 59.9–60.8 |
| 22 | Sm. 19.7–20.5 | 77 | Sm. 60.8–61.5 |
| 23 | Sm. 20.5–21.2 | 78 | Sm. 61.5–62.3 |
| 24 | Sm. 21.2–21.10 | 79 | Sm. 62.3–63.2 |
| 25 | Sm. 21.10–22.6 | 80 | Sm. 63.2–64.3 |
| 26 | Sm. 22.6–23.1 | 81 | Sm. 64.3–65.1 |
| 27 | Sm. 23.1–23.8 | 82 | Sm. 65.1–65.8 |
| 28 | Sm. 23.8–23.15 | 83 | Sm. 65.8–66.7 |
| 29 | Sm. 23.15–24.5 | 84 | Sm. 66.7–66.14 |
| 30 | Sm. 24.5–24.8 | 85 | Sm. 66.14–67.7 |
| 31 | Sm. 24.8–25.5 | 86 | Sm. 67.7–68.3 |
| 32 | Sm. 25.5–25.9 | 87 | Sm. 68.3–69.2 |
| 33 | Sm. 25.9–26.5 | 88 | Sm. 69.2–70.1 |
| 34 | Sm. 26.5–26.10 | 89 | Sm. 70.1–70.8 |
| 35 | Sm. 26.10–27.2 | 90 | Sm. 70.8–71.6 |
| 36 | Sm. 26.10–27.9 | 91 | Sm. 71.6–71.14 |
| 37 | Sm. 27.9–28.3 | 92 | Sm. 71.14–72.5 |
| 38 | Sm. 28.3–28.10 | 93 | Sm. 72.5–73.2 |
| 39 | Sm. 28.10–29.3 | 94 | Sm. 73.2–73.9 |
| 40 | Sm. 29.3–30.1 | 95 | Sm. 73.9–74.7 |
| 41 | Sm. 30.1–30.8 | 96 | Sm. 74.7–75.3 |
| 42 | Sm. 30.8–31.2 | 97 | Sm. 75.3–75.11 |
| 43 | Sm. 31.2–31.9 | 98 | Sm. 75.11–76.8 |
| 44 | Sm. 31.9–32.6 | 99 | Sm. 76.8–77.4 |
| 45 | Sm. 32.6–33.3 | 100 | Sm. 77.4–78.5 |
| 46 | Sm. 33.3–33.10 | 101 | Sm. 78.5–79.4 |
| 47 | Sm. 33.10–34.1 | 102 | Sm. 79.4–80.4 |
| 48 | Sm. 34.1–35.3 | 103 | Sm. 80.4–81.3 |
| 49 | Sm. 35.3–36.2 | 104 | Sm. 81.3–81.9 |
| 50 | Sm. 36.2–37.2 | 105 | Sm. 81.9–82.4 |
| 51 | Sm. 37.2–38.2 | 106 | Sm. 82.4–83.4 |
| 52 | Sm. 38.2–39.4 | 107 | Sm. 83.4–84.4 |
| 53 | Sm. 39.4–40.3 | 108 | Sm. 84.4–85.4 |
| 54 | Sm. 40.3–41.4 | 109 | Sm. 85.4–85.11 |
| 55 | Sm. 41.4–42.6 | | |

Table A.4.11: Description of sample contents (1,500 $w$) for Bernard of Clairvaux's *Sermones super Cantica canticorum* in figs. 4.3–4.4.

| index | *SBO* | index | *SBO* |
|---|---|---|---|
| 1 | Ep. 1 (196: 1593a-1594b); Ep. 2 (196: 1594b-1596a); Ep. 3 (196: 1596b-1597b) | 8 | Ep. 31 (196: 1622c-1622d); Ep. 32 (196: 1623a-1623c); Ep. 33 (196: 1623c-1625c); Ep. 34 (196: 1625d-1626c) |
| 2 | Ep. 4 (196: 1597b-1598c); Ep. 5 (196: 1598d-1600a); Ep. 6 (196: 1600b-1601b); Ep. 7 (196: 1601c-1601d) | 9 | Ep. 35 (196: 1626d-1631a); Ep. 36 (196: 1631b-1631c) |
| 3 | Ep. 7 (196: 1601d-1603a); Ep. 8 (196: 1603b-1605a); Ep. 9 (196: 1605b-1605d) | 10 | Ep. 36 (196: 1631c-1632c); Ep. 38 (196: 1632c-1635b) |
| 4 | Ep. 10 (196: 1606a-1607d); Ep. 11 (196: 1608a-1608c); Ep. 12 (196: 1608c-1609a); Ep. 15 (196: 1609b-1610a) | 11 | Ep. 38 (196: 1635b- 1636c); Ep. 40 (196: 1636d-1639a) |
| 5 | Ep. 15 (196: 1610a-1610c); Ep. 16 (196: 1610d-1613c); Ep. 17 (196: 1613d-1614a) | 12 | Ep. 40 (196: 1639b-1639d); Ep. 41 (196: 1640a-1640b); Ep. 42 (196: 1640c-1641c); Ep. 43 (196: 1641c-1643b) |
| 6 | Ep. 17 (196: 1614a-1616a); Ep. 18 (196: 1616b-1617c); Ep. 19 (196: 1617d-1618a) | 13 | Ep. 43 (196: 1643b-1644a); Ep. 44 (196: 1644a-1645a); Ep. 45 (196: 1645b-1646d); Ep. 46 (196: 1647a-1647c) |
| 7 | Ep. 23 (196: 1618c-1619a); Ep. 27 (196: 1619c-1620a); Ep. 29 (196: 1620b-1621c); Ep. 31 (196: 1621d-1622c) | 14 | Ep. 46 (196: 1647d- 1648c); Ep. 47 (196: 1648d-1649a); Ep. 50(196: 1649c-1650c); Ep. 51 (196: 1651a-1651d) |
| 1 | Sm. 69 (144: 897c-901c) | 14 | Sm. 62 (144: 856a-857c); Sm. 23 (144: 627b-629b) |
| 2 | Sm. 69 (144: 901d-902b); Sm. 43 (144: 732b-735b) | 15 | Sm. 23 (144: 629c-633b) |
| 3 | Sm. 43 (144: 735c-736b); Sm. 55 (144: 811c-814c) | 16 | Sm. 23 (144: 633c-637a); Sm. 27 (144: 649a-649b) |
| 4 | Sm. 55 (144: 814d-815c); Sm. 56 (144: 815d-818b) | 17 | Sm. 27 (144: 649c-652c); 9.42. hom. (144: 548c-549b) |
| 5 | Sm. 56 (144: 818c-822b) | 18 | 9.42. hom. (144: 549c-553a) |
| 6 | Sm. 56 (144: 822b-822d); Sm. 58 (144: 828d-832a) | 19 | 9.42. hom. (144: 553b); Sm. 29 (144: 660b-663d) |
| 7 | Sm. 58 (144: 832b-834c); Sm. 59 (144: 834d-836b) | 20 | Sm. 29 (144: 664a-666a); Sm. 26 (144: 646b-647d) |
| 8 | Sm. 59 (144: 836c-838d); Sm. 11 (144: 557a-558a) | 21 | Sm. 26 (144: 648a-649a); Sm. 40 (144: 717a-719d) |
| 9 | Sm. 11 (144: 558b-561d) | 22 | Sm. 40 (144: 720a-722c); Sm. 44 (144: 736b-737b) |
| 10 | Sm. 11 (144: 562a-563a); Sm. 60 (144: 839b-841d) | 23 | Sm. 44 (144: 737c-740d); Sm. 47 (144: 761c-761d) |
| 11 | Sm. 60 (144: 842a-845d) | 24 | Sm. 47 (144: 762a-765c) |
| 12 | Sm. 60 (144: 846a); Sm. anonym. (144: 848b-851d) | 25 | Sm. 47 (144: 766a-766b); Sm. 52 (144: 794b-797b) |
| 13 | Sm. anonym. (144: 852a-853b); Sm. 62 (144: 853b-855d) | | |

Table A.4.12: Description of sample contents (1,500 *w*) for Nicholas's sermons and letters in figs. 4.3–4.4.

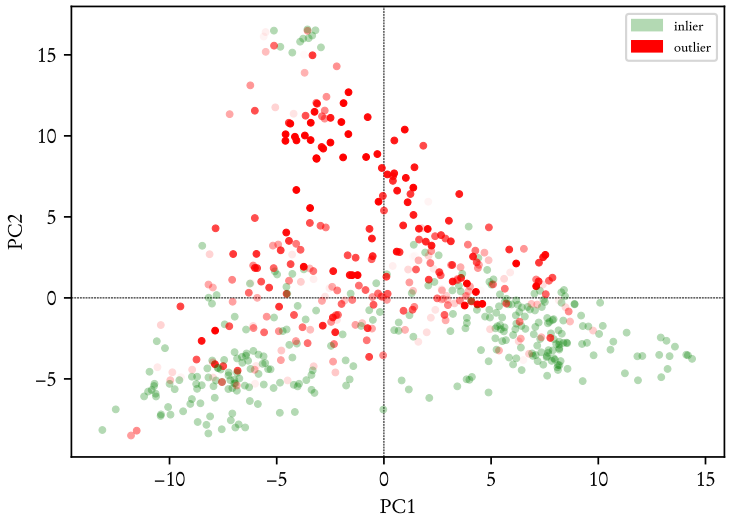



Figure A.20: Rehearsal of fig. 5.2 on p. 158. Intuition of outlier detection for Elisabeth of Schönau's visions. The intensity of red indicates the degree of 'anomaly.' As becomes clear from the plot, inliers are not necessarily located at the outer edges of a population. Settings: $s$–$l$ = 500 | *type* = most-frequent words | $n$ = 1,000 | *vect.* = standard-scaled tfidf-weighted frequencies | *expl. var.* = 6.29%.

# A.5 Addenda to Chapter 5

## A.5.1 The LOF algorithm.

Fig. A.20 visualizes the window samples of Elisabeth of Schönau's visions as data points in a two-dimensional space (PCA), and visualizes the attributions made by a local outlier detection algorithm (LOF). This algorithm combines the intuition of $k$ nearest neighbours and outlier detection.[90] The consecutive steps were as follows:

- Initially, a feature subset of the 1,000 MFW from the entire corpus vocabulary is produced. Once all of the windows have been transformed to standard-scaled tfidf-vectors with respect to this subset, the LOF algorithm is summoned to distinguish inliers from outliers.
- LOF makes the distinction between text samples as either inliers or outliers by a boolean decision function. The number of neighbours was set to $k$=20.
- LOF makes its decision by adhering to a fixed proportion of outliers. This threshold, which is sometimes called the 'contamination' parameter, and which functions in a way comparable to a classification threshold (see p. 345), allows the analyst to decide at which point a sample is either 'in' or 'out.' However, choosing only one threshold with a strict in- and exclusion of text samples risks a unilateral take on the data's dynamics. Instead, different thresholds were searched and their decision outputs collected and combined. This yielded

[90] Markus M. Breunig et al., "LOF: Identifying Density-Based Local Outliers," *ACM SIGMOD Record* 29, no. 2 (2000): 93–104.

a more layered and nuanced classification of outliers, which does not make hard classifications, but finds consensus over different calculations of a sample's 'outlier-ness.' This degree is visualized by the intensity of red in fig. 5.2. Hard reds are undeniable anomalies within Elisabeth's visions. The softer the red appears, the less outspoken that anomaly becomes. The greens in the PCA plot are at the core of the distribution's 'normal' behaviour, and are not considered outliers

# A.6 Addenda to Chapter 6

Table A.6.1: Word correspondences between *Mulierem inquiunt* (*Vita Hildegardis*, 3.26, ll. 15–38, 68–9) and Hildegard's canonical works. Latin text of the *vita* in left column edited by Monika Klaes.

| ***Mulierem inquiunt*** (full text) | **Hildegardian phrasing** |
|---|---|
| Mulierem inquiunt quandam acriter a | |
| demonio muto uexatam, **super quam** | ... ut sol rem aliquam calefacit **super quam**... (*Scivias*, *Protestificatio* l. 29) |
| et fratres de Lacu plurimum laboraverant, | |
| **cum** ad se **magno** labore uirorum in | ... **cum magno** sacramento incarnationis sue (*Scivias*, II.3.34, l. 737) ... |
| lecto deportata esset, pia mater | |
| **audacie et presumptioni** demonis | ... **audatie et presumptionis** temeritatem... |
| uerbis a Spiritu sancto | (*LDO,* I.3.14 l. 1) |
| prolatis \| confidenter resistens ab orationibus et | |
| benedictionibus **non cessauit, quousque** | ... **nec** in hoc **cessabit, quousque** numerus |
| per gratiam | ... (*LDO,* III.4.9, l. 10) |
| Dei ab hoste maligno eam liberauit. | |
| Simili modo et alia mulier, | |
| **que propter** furorem insanie diris | ... Homo autem, **qui propter** timorem ... (*LDO*, I.2.35, l. 24) |
| uinculis ligata fuit, cum ad se | |
| adducta fuisset, solui eam monuit, et | |
| statim sub admiratione omnium qui | |
| aderant sanitate | |
| mentis et corporis recepta cum gratiarum | |
| actione remeauit ad propria. | |
| Item de claustro Schefeneburch | |
| sororem quandam diabolus | |
| ad sancta opera, | |
| **orationes, uigilias et ieiunia,** | ... que **cum uigiliis et ieiuniis** ac **orationibus**. ... (*LDO*, II.1.39.2) |
| **ad perceptionem** quoque | ... **ad percipiendum** idem |
| **sacramentorum** | **sacramentum** ... (*Scivias*, II.6, l. 243) |
| simulans se esse angelum lucis | |
| hortabatur et criminalium confessione, | |
| quibus numquam subiacuerat | |
| confundere eam nitebatur. | |
| Inter **que etiam ita** eam | ... ager fructum proferens **qui etiam ita** Deo est consecratus ... (*Scivias*, II. 5.48, l. 1494) |
| **afflixit, quod** nomina et | ... in tantum **affligitur, quod** ... (*LDO*, I.4.64, l. 27) |
| aspectum **quorundam hominum** et | ... cerebrum **quorundam hominum** |
| animalium | igneum et siccum est ... (*LDO*, I. 2.32, l. 166) |
| **in tantum** abhorrebat, **quod** ipsis uisis | ... **in tantum** affligitur, **quod** ... (*LDO*, I.4.64, |
| uel auditis horribili uoce per | l. 27) |
| longam horam perstrepebat. | |
| Hec a priore et conuentu | |
| cum litteris ad sanctam uirginem | |
| missa ab ea et confortata et a | *confortare* is an often used word in Hildegard's oeuvre. |
| diaboli errore est liberata. | |
| Eadem uirtute alia duas | |

Table A.6.1: Word correspondences between *Mulierem inquiunt* and Hildegard's canonical works (continued).

| ***Mulierem inquiunt*** (full text) | **Hildegardian phrasing** |
|---|---|
| mulieres obsessas a<br>demonio liberauit, quarum una,<br>cum esset **paupercula** et ceca, | ... Ego igitur **paupercula** et inbecillis ...<br>(*LDO*, *Prologus*, l. 27) |
| in elemosinam eius recepta in spirituali<br>habitu uitam **feliciter consummauit** | sed tamen ea bono fine<br>**feliciter consummauit** (*Vita sancti Disibodi*, 10, l. 136, p. 63) |

Table A.6.2: Word correspondences between *Cum beata* (*Vita Hildegardis*, 3.27, ll. 4–35, 69–70) and Hildegard's canonical works. Latin text of the *vita* in left column edited by Monika Klaes.

| ***Cum beata*** (full text) | **Hildegardian phrasing** |
|---|---|
| "Cum beata inquiunt mater Domino multis<br>**laborum** certaminibus deuote militasset,<br>uite presentis **tedio affecta** | ... **labore** et **tedio affectum** ...<br>(*LDO,* I.2.35, l. 12) |
| **dissolui et esse cum Christo**<br>cottidie **cupiebat**. Cuius<br>**desiderium** Deus exaudiens finem suum,<br>sicut ipsa preoptauerat, **spiritu prophetie** | ... in candore tantum celestis **desiderii**<br>**querit dissolui et esse cum Christo** ...<br>(*Scivias*, III. 10.22 l. 648)<br>... in **spiritu prophetiae** cognouerunt ...<br>(*LVM*, 2.30, l. 545) |
| ei reuelauit, quem et<br>sororibus predixit. Aliquamdiu<br>**itaque infirmitate laborans** octogesimo<br>secundo **etatis sue anno** | **itaque** ... **infirmitate laborasset**,<br>uicesimo **etatis sue anno** ... (*Vita Sancti Ruperti*, 11, l. 358, p. 103.) |
| quinto decimo Kalendarum-<br>octobris ad sponsum celestem felici<br>transitu migrauit. Filie autem ipsius,<br>quarum omne gaudium et solacium in<br>ipsa erat, funeri<br>dilecte matris amarissime flentes<br>assistebant. Nam licet de premiis eius et<br>de suffragiis sibi per ipsam<br>conferendis non dubitarent, | |
| **propter** discessionem tamen eius, | ... **propter** sanguinem agni, **per quem** ...<br>(*LDO*, III.5.37, l. 25) |
| **per quam semper** consolabantur, | ... pugna, **per quam semper** ...<br>(*Scivias*, III.6.30, l. 750) |
| maximo cordis merore afficiebantur.<br>Deus uero, cuius meriti apud<br>se esset in transitu suo euidenter<br>declarauit.\| Nam supra habitaculum, in<br>quo sancta uirgo<br>primo crepusculo noctis dominice diei<br>felicem animam Deo reddidit, duo<br>lucidissimi et **diuersi coloris** arcus | ... et **diuerso colore** depicta ...<br>(*Scivias*, III.10.9, l. 433) |
| **in firmamento** apparuerunt, qui ad | ... magna **in firmamento** discurrunt, sic ...<br>(*LDO*, I.4.51, l. 14) |
| magnitudinem magne platee **se**<br>**dilatauerunt**<br>**in quatuor partes terre se**<br>**extendentes**, | ... **se dilatauerant**, in hac palude ...<br>(*LVM*, I.121, l. 1852)<br>... in quattuor partes se diuiserunt...<br>(*Scivias*, II.7, l. 84) |

Table A.6.2: Word correspondences between *Cum beata* and Hildegard's canonical works (continued).

| ***Cum beata*** (full text) | **Hildegardian phrasing** |
|---|---|
| | ... ad **quattuor plagas orbis extendentem**, ... (*Scivias*, I.4.9, l. 398) |
| | ... super **quatuor partes terrae** ... (*LDO* II.1.8, l. 65) |
| quorum alter **ab aquilone ad austrum**, | ... **ad austrum et altera ad aquilonem** ... (*LDO*, I.4.49, l. 66) |
| alter ab **oriente ad** occidentem **procedebant**. | ... **ad orientem procedebant** rami a ... (*Scivias*, III.4, l. 59) |
| | ... **procedebant**, se in altitudine ... (*Scivias*, III.6.35, l. 1010) |
| At **in summitate, ubi** hi | ... **in** cuius **summitate, ubi** locus ... (*LDO*, III.4.1, l. 10) |
| duo arcus iungebantur, clara lux ad quantitatem lunaris circuli emergebat, que late se protendens **tenebras noctis** ab habitaculo depellere **uidebatur. In** hac | ... **tenebras noctis** cum mala ... (*LVM*, III.28, l. 527) ... excellere **uideretur; in** quo ... (*LDO*, III.1.1, l. 5) |
| luce crux rutilans uisa est, primum parua, sed crescendo postea immensa, circa **quam innumerabiles uarii** | ... **qui innumerabiles** in numero ... (*LVM*, I.49, l. 777) ... **uarii coloris** induta est, ... (*LVM*, 2.47, l. 935) |
| **coloris circuli, in quibus** singulis singule rutilantes crucicule oriebantur, cum circulis suis crescentes priore t amen minores conspiciebantur. | ... **circuli, in quo** similitudo ... (*LDO*, I.2.1, l. 160) |
| Et cum he **in firmamento se** | ... **se in firmamento** distantes signati ... (*LDO*, I.4.22, l.1) |
| dilatassent, **latitudine sua** | ... **latitudinem** multam habentem, ac ... (*LVM*, IV.50, l. 1090) |
| **ad orientem** magis pertingebant et **ad terram uersus** domum in | ... et **ad orientem uersam** ... (*LDO*, III.5.2, l.19) |
| qua sancta uirgo transierat, declinare uise **totum montem clarificabant.** | ... **totum mundum illuminaret** ... (*LDO*, III.5.9, l. 100) |
| Et credendum quod **hoc signo** Deus | ... **signis se ostendunt**; quoniam ... (*LDO*, III.1.5, l. 4) |
| **ostendit, quanta claritate** dilectam suam in celestibus illustrauerit." | ... **tanta claritate** fulgebat, ut ... (*LDO*, III.3.1, l. 7) |

# A.7 Addenda to Chapter 7

Table A.7.1: Word correspondences between the donation charter of Charlemagne to Saint-Denis (D Kar 286) and the anonymous *Pseudo-Turpin*.

| **D Kar 286** | ***Pseudo-Turpin*** |
|---|---|
| Igitur ob id ad honorem dei et tantorum dominorum nostrorum reverentiam tale fieri decrevimus **praeceptum**, quod volumus imperpetuum esse mansurum, videlicet quod **omnes Franciae reges**, omnes archiepiscopi et **episcopi** tam **praesentes** quam et **futuri** ob amorem domini dei et nostri salvatoris Iesu **Christi** honorem ac reverentiam deferant venerandae matri **ecclesiae** domni | Et **praecepit** ut<br><br>**omnes Franciae reges**<br>et **episcopi, praesentes** et **futuri**,<br><br>pastori eiusdem **ecclesiae**<br>essent obedientes in **Christo.** |
| Franciae **reges** alibi quam in ecclesia saepe fati domni Dionysii sint **coronati** nec archiepiscopi et **episcopi** confirmati aut ad sacram beati Petri sedem **recepti** et **damnati** absque assensu et consilio abbatis. | Nec **reges** sine eius consilio<br>essent **coronati,**<br>nec **episcopi** ordinati,<br>nec apud Romam **recepti** essent aut **dampnati.** |
| Post vero multa **eidem ecclesiae** bona per nos oblata ac concessa privilegia ego Karolus Francorum rex deposito de capite meo regni diademate et sanctorum martyrum altari superposito talia cunctis qui aderant audientibus dixi: Sanctissime domine Dionysi hiis regni Franciae regiis insigniis et ornamentis libenter me spolio, ut deinceps eius regale habeas, teneas atque possideas dominium et in signum rei **quatuor** | Rursum **post** plurima dona **eidem ecclesiae** in<br><br><br><br><br><br><br>praedio dedit praecipiens ut unusquisque possessor uniuscuiusque domus totius Galliae **quattuor num-mos** |
| modo **aureos** tibi offero **bizancios**, ut omnes tam praesentes quam et futuri sciant et agnoscant, quod a deo solo et a te regnum Franciae teneo tuoque ac tuorum sociorum fretus auxilio et suffragantibus meritis illud ancipiti gladio defendo obsecrans atque obtestans omnes successores nostros reges, ut **annuatim simile** faciant et in oblatione submittendo ac tangendo caput illos quatuor supradictos bysancios offerant non proinde astricti humanae servituti, sed potius divinae, quae summa libertas appellari debet, quippe deo servire sit regnare, sed etiam omnes regni nostri proceres et obtimates pro qualicumque domo sua eidem ecclesiae memoratae **quatuor singulis** annis aureos persolvant **nummos pro** illius **augmento** ab **aedificio** Dagoberti regis excellentissimi usque ad crucifixum; similiter **omnes** homines **servituti** addictos emancipans et eos imperpetuum faciens **liberos** tam praesentes quam et futuros, **qui libenter eosdem quatuor** aureos **nummos** | **annuatim** ad<br><br><br><br><br><br>

Table A.7.1: Word correspondences between the donation charter of Charlemagne to Saint-Denis (D Kar 286) and the anonymous *Pseudo-Turpin* (continued).

| **D Kar 286** | ***Pseudo-Turpin*** |
|---|---|
| | Christianis similiter qui propria sua pro divino amore dimiserant et in Hyspania in bellis Sarracenorum martirii coronam acceperant. Iccirco nocte proxima regi dormienti beatus Dionisius apparuit, eumque excitavit, dicens ei: Illis qui tua |
| daturi sunt, | ammonitione et exemplo tuae probitatis animati in bellis Sarracenorum in Hyspania mortui et morituri sunt, delictorum omnium suorum veniam, et illis qui **nummos ad haedificandum ecclesiam** meam dant et **daturi sunt**, gravioris sui vulneris |
| quos beati **Dionysii Francos** proinde **vocari** volo et appellari iubeo. | medicinam a Deo impetravi.<br>His a rege relatis, populi nummos saluberrimae promissionis suae libenter ex more dabant, et qui libentius reddebat **Francus** sancti **Dionisii** ubique vocabatur, quoniam liber ab omni servitute, rege praecipiente, erat. Hinc mos surrexit ut terra ilia quae antea **vocabatur** Gallia, nunc vocatur Francia, id est, ab omni servitute aliarum gentium libera. Quapropter Francus liber dicitur, quia super omnes gentes alias decus et dominatio illi debetur. |

# A.8 Addenda to Chapter 8

Table A.8.1: Table of function words used for generating fig. 8.2b.

| 1–50 | 51–100 | 101–150 | 151–200 | 201–250 | 251–300 | 301–350 |
|---|---|---|---|---|---|---|
| et | eum | ipsis | uestra | has | quandam | alius |
| in | his | contra | tuam | quisquis | aliquem | ultra |
| non | ei | numquam | tui | utique | huc | utpote |
| est | illud | secundum | aliquo | qualis | mearum | illarum |
| ad | nisi | supra | meis | hos | deinde | neuter |
| ut | qua | hac | hunc | quocumque | intra | quoad |
| que | ipsum | modo | illos | rursum | ulla | proinde |
| de | ac | eas | quidam | illuc | nondum | etsi |
| quam | eam | illam | postea | quosdam | aliquibus | quodlibet |
| cum | ubi | illis | meo | tandem | quendam | qualibet |
| quod | scilicet | ante | ideo | semel | rursus | prope |
| qui | tamen | tua | diu | quamuis | uersus | uestros |
| si | tunc | aliquid | prout | tuas | quot | quoslibet |
| uel | ipso | mea | namque | talibus | quantus | utramque |
| etiam | earum | ipsam | extra | tum | quales | qualiter |
| hoc | apud | illo | aliud | aque | queque | quantumcumque |
| sed | ille | idem | quibusdam | plerisque | utriusque | unusquisque |
| autem | huius | usque | sola | aliquod | cuiusdam | qualem |
| nec | itaque | unus | quicumque | meos | rem | aliqui |
| eius | hanc | ergo | meus | talem | uestrum | istud |
| per | eorum | item | quidquid | utrum | at | quascumque |
| ab | super | tu | iis | donec | nuper | tuos |
| quo | ego | denique | aduersum | ue | aliquos | quisquam |
| tam | post | aliqua | simul | quacumque | plerumque | propterea |
| quoque | solum | quodam | erga | quemadmodum | harum | quasdam |
| enim | illa | illius | eadem | utrumque | quibuscumque | cuiuslibet |
| sicut | nunc | quoniam | ipsorum | quatinus | alicuius | demum |
| se | uos | tuum | huic | uti | idcirco | hucusque |
| ex | sine | nullum | uestri | tuus | ipsarum | quicquam |
| quidem | ibi | unum | modum | mox | relique | tuorum |
| uero | tantum | postmodum | meam | totus | absque | utroque |
| ipse | eo | alia | horum | utraque | quodcumque | alicubi |
| aut | inter | sub | aliquam | alter | he | suus |
| nos | quantum | inde | coram | ob | quare | talia |
| atque | nulla | tamquam | aduersus | quandoque | procul | quamquam |
| quid | ipsius | quomodo | aliquando | ubique | uelut | plerosque |
| pro | hic | nam | utinam | alterum | quilibet | illic |
| quasi | hinc | quadam | uestram | etenim | is | aliquas |
| quia | eos | una | illorum | pleraque | talis | necnon |
| eis | immo | meum | quinque | recte | quisque | tuarum |
| ipsa | iuxta | illum | tuo | quoddam | siquidem | quousque |
| sic | propter | solus | ipsas | quotiens | quamcumque | interdum |
| quanto | ipsi | neque | umquam | hi | quantacumque | aliquarum |
| ita | illi | quando | olim | tantus | cuiuscumque | unicuique |
| id | seu | noster | quecumque | uester | quibuslibet | utrique |
| unde | ea | licet | nullus | tot | meas | cuilibet |
| iam | adhuc | primum | circa | meorum | ubicumque | qualibus |
| quis | siue | tuis | tales | altera | quolibet | aliquorum |
| quippe | dum | ipsos | iste | tali | sin | aduersa |
| uidelicet | mei | ista | illas | uestris | quidquam | fas |

Table A.8.2: Table of function words used for generating fig. 8.5b.

| 1–50 | 51–100 | 101–150 | 151–200 | 201–250 | 251–300 | 301–350 |
|---|---|---|---|---|---|---|
| et | quis | ergo | tamquam | has | utriusque | tuarum |
| in | his | meo | quadam | aduersus | altera | quousque |
| non | quippe | iuxta | aliud | iste | alter | is |
| est | uidelicet | illi | tuus | utique | quandoque | quisque |
| ut | eum | secundum | meis | uestram | ubique | quamcumque |
| ad | qua | ipsi | primum | qualis | plerumque | propterea |
| que | ei | adhuc | illum | illorum | etenim | quantacumque |
| de | tu | modo | ista | quinque | prorsus | cuiuscumque |
| quod | ac | supra | meam | ipsas | pleraque | tuos |
| quam | illud | seu | simul | umquam | idcirco | quibuslibet |
| cum | tamen | tuum | ipsos | quecumque | recte | nusquam |
| qui | ipsum | siue | diu | ulla | quamdiu | quolibet |
| si | ille | hac | erga | quisquis | quoddam | sin |
| uel | ubi | contra | aliquo | quocumque | intra | istud |
| etiam | scilicet | ipsis | uestra | alterum | quotiens | alius |
| sed | eam | idem | hunc | ob | hi | quidquam |
| hoc | sine | aliquid | quidam | illas | tantus | utpote |
| nec | nunc | tuo | ideo | tali | quantus | uerumtamen |
| autem | super | eas | illos | tandem | uester | neuter |
| eius | tunc | tuis | postea | semel | qualiter | quicquam |
| per | apud | illam | eadem | uti | uestris | quoad |
| tam | tua | ante | quot | talis | quandam | proinde |
| quo | ipso | illis | namque | rursum | aliquem | etsi |
| enim | illa | nam | extra | illuc | huc | tuorum |
| se | eorum | usque | prout | talibus | siquidem | quodlibet |
| ab | huius | aliqua | tuas | ultra | mearum | prope |
| quoque | earum | tuam | quibusdam | quosdam | deinde | quoslibet |
| sicut | itaque | illo | quidquid | meas | nuper | qualibet |
| ex | post | ipsam | utinam | talem | absque | uestros |
| quidem | hanc | unus | quicumque | tum | alicuius | cuique |
| quia | numquam | denique | horum | aque | aliquibus | uterque |
| ipse | solum | unum | circa | quacumque | quendam | utramque |
| uero | inter | item | iis | plerisque | rursus | fas |
| quid | uos | neque | huic | demum | procul | quantumcumque |
| nos | tantum | nullum | aduersum | aliquod | uelut | utcumque |
| aut | quantum | illius | quamuis | rem | quilibet | unusquisque |
| atque | nulla | inde | modum | meos | quales | qualem |
| ita | ibi | quomodo | nullus | utrum | cuiusdam | aliqui |
| ego | mei | sola | ipsorum | at | aliquos | quascumque |
| sic | eo | quodam | totus | hucusque | uestrum | aliquis |
| pro | tui | quoniam | aliquam | ue | harum | quisquam |
| mea | meum | licet | olim | queque | quibuscumque | infra |
| nisi | dum | una | uestri | quemadmodum | ipsarum | quasdam |
| quasi | hic | quando | hos | nondum | ubicumque | cuiuslibet |
| unde | ipsius | sub | aliquando | quatinus | relique | tale |
| ipsa | immo | postmodum | tales | utrumque | illarum | cuiusque |
| eis | propter | meus | coram | uersus | quodcumque | ullus |
| quanto | eos | noster | tot | mox | interdum | penes |
| id | hinc | solus | mi | utraque | quare | sane |
| iam | ea | alia | donec | meorum | he | utroque |

# Primary sources


Anselmus Cantuariensis. *Cur deus homo*. In *S. Anselmi Cantuariensis archiepiscopi opera omnia*, edited by Franciscus Salesius Schmitt, 2:39–133. Apud Thomam Nelson et Filios, 1946.

Baldricus Dolensis. *Poèmes*. Edited and translated by Jean-Yves Tilliette. Auteurs latins du moyen âge. Paris: Les Belles Lettres, 1998–2002.

Bernardus Claraevallensis. *Sämtliche Werke*. Edited by Gerhard B. Winkler. Innsbruck: Tyrolia-Verlag, 1990–9.

———. *Sancti Bernardi Opera*. Edited by Jean Leclercq, Henri M. Rochais, and Charles Holwell Talbot. Rome: Editiones Cistercienses, 1957–77.

Bruder, Petro, ed. *Acta inquisitionis de virtutibus et miraculis S. Hildegardis, Magistrae sororum ord. S. Benedicti in monte S. Ruperti juxta bingium ad rhenum. Ex originali archetypo transcripsit notisque illustravit.* 116–29. *Analecta Bollandiana* 2. 1883.

Burgess, Glyn S., and Anne Elizabeth Cobby, eds. and trans. *The Pilgrimage of Charlemagne (Le Pèlerinage de Charlemagne); Aucassin and Nicolette (Aucassin et Nicolette)*. Garland Library of Medieval Literature, A:47. New York: Garland Publishing, 1988.

Dufour, Jean, ed. *Recueil des actes de Louis VI, roi de France (1108-1137)*. Vol. 3. Chartes et diplômes relatifs à l'histoire de France. Paris: Diffusion de Bocard, 1993.

Eadmerus Cantuariensis. *The Life of St. Anselm, Archbishop of Canterbury*. Edited and translated by Richard William Southern. Medieval Texts. London: Nelson, 1962.

Einhardus. *Einhard. Vita Karoli Magni*. Edited by Ludwig Gompf. Münster: Verlag Aschendorff, 1971.

Ekbertus Schoenaugiensis. *De obitu Elisabeth*. In *Die Visionen der hl. Elisabeth und die Schriften der Aebte Ekbert und Emecho von Schönau*, edited by Ferdinand W.E. Roth, 263–78. Brünn: Raigerner Benediktiner Buchdruckerei. Verlag der "Studien aus dem Benediktiner- und Cistercienser-Orden", 1884.

Elisabeth Schoenaugiensis. *Die Visionen der hl. Elisabeth und die Schriften der Aebte Ekbert und Emecho von Schönau*. Edited by Ferdinand W.E. Roth. Brünn: Raigerner Benediktiner Buchdruckerei. Verlag der "Studien aus dem Benediktiner- und Cistercienser-Orden", 1884.

———. *Elisabeth of Schönau: The Complete Works*. Translated by Anne L. Clark and Barbara Newman. The Classics of Western Spirituality. New York: Paulist Press, 2000.

———. "Revelationes B. Elisabeth. Disquisitio critica una cum textibus latino et catalaunensi." Edited by Livarius Oliger. *Antonianum* 1 (1926): 24–83.

Gaufridus Autissiodorensis. *Goffredo di Auxerre: Super Apocalypsim*. Edited by Ferruccio Gastaldelli. Temi e Testi 17. Rome: Storia e Letteratura, 1970.

———. *Libri III–VI Sancti Bernardi abbatis Claraevallensis vita et res gestae*. In *PL* 185:301–51, edited by Jean Mabillon.

Godefridus Sancti Disibodi and Theodericus Epternacensis. *The Life of Hildegard*. In *Jutta and Hildegard: The Biographical Sources*, edited and translated by Anna Silvas, 118–210. Medieval Women: Texts and Contexts 1. Turnhout: Brepols, 1998.

———. *Vita sanctae Hildegardis virginis*. Edited by Monika Klaes. *CC CM* 126. Turnhout: Brepols, 1993.

Guibertus de Novigento. *Autobiographie*. Edited and translated by Edmond-Réné Labande. Les classiques de l'histoire de France au Moyen Age 34. Paris: Les Belles Lettres, 1981.

———. *De sanctis et eorum pigneribus*. In *Quo ordine sermo fieri debeat; De bucella Iudae data et de veritate dominici corporis; De sanctis et eorum pigneribus*, edited and translated by Robert B.C. Huygens, 79–175. *CC CM* 127. Turnhout: Brepols, 1993.

———. *Guibert de Nogent:* Dei gesta per Francos *et cinq autres textes*. Edited by Robert B.C. Huygens. *CC CM* 127A. Turnhout: Brepols, 1996.

———. Monodies *and* On the Relics of Saints. *The Autobiography and a Manifesto of a French Monk from the Time of the Crusades*. Translated by Joseph McAlhany and Jay Rubenstein. New York: Penguin Books, 2011.

———. *Tropologiae in prophetas Osee et Amos ac lamentationes Jeremiae*. In *PL* 156:337–488, edited by Luc d'Achery.

Guibertus Gemblacensis. *De combustione monasterii Gemblacensis*. Edited by Georg Heinrich Pertz, 563–4. *MGH Scriptores* 8. Hannover, 1848.

———. *Guiberti Gemblacensis Epistolae quae in codice B.R. BRUX. 5527–5534 inueniuntur*. Edited by Albert Derolez. *CC CM* 66–66A. Turnhout: Brepols, 1988–9.

———. *Vita sanctae Hildegardis retractata*. In *Vita sanctae Hildegardis*, edited by Monika Klaes, 93–106. *CC CM* 126. Turnhout: Brepols, 1993.

Guillelmus de Sancto Theodorico. *Disputatio adversus Petrum Abaelardum*. In *Guillelmi a Sancto Theodorico opera omnia*, edited by Paul Verdeyen, 5:13–59. *CC CM* 89A. Turnhout: Brepols, 2007.

———. *Liber I Sancti Bernardi abbatis Claraevallensis vita et res gestae*. In *PL* 185:225–68, edited by Jean Mabillon.

Guillelmus de Sancto Theodorico, Arnaldus Bonaevallensis, and Gaufridus Autissiodorensis. *First Life of Bernard of Clairvaux*. Edited and translated by Hilary Costello. Cistercian Fathers Series 76. Collegeville, MN: Liturgical Press, 2015.

———. *Vita prima Sancti Bernardi Claraevallis abbatis*. In *Guillelmi a Sancto Theodorico opera omnia*, edited by Paul Verdeyen, 6:13–233. *CC CM* 89B. Turnhout: Brepols, 2011.

Guillelmus Sancti Dionysii. *Epistola Willelmi monachi S. Dionysii ad quosdam ex suis comonachis*. In "Quelques mots à propos d'une lettre de Guillaume de Saint-Denis", edited by Edmond-Réné Labande, 23–35. *Mélanges offerts à Rita Lejeune* 1. Gembloux: Éditions J. Duculot, 1969.

———. *Vie de Suger par le moine Guillaume*. In *Oeuvres / Suger*, edited and translated by Françoise Gasparri, 2:292–357. Les classiques de l'histoire de France au Moyen Age 41. Paris: Les Belles Lettres, 2001.

Hämel, Adalbert, and André de Mandach, eds. *Der Pseudo-Turpin von Compostela*. München: Verlag der Bayerischen Akademie der Wissenschaften, 1965.

Herrad Hohenburgensis. *Hortus deliciarium*. Edited by Rosalie Green, Michael Evans, Christine Bischoff, and Michael Curschmann. Studies of the Warburg Institute 36. Leiden: Brill, 1979.

Hildegardis Bingensis. *Epistolarium Pars Prima I–XC*. Edited by Lieven Van Acker. *CC CM* 91. Turnhout: Brepols, 1991.

Hildegardis Bingensis. *Epistolarium Pars Secunda XCI–CCL*$^{R}$. Edited by Lieven Van Acker. *CC CM* 91A. Turnhout: Brepols, 1993.

———. *Epistolarium Pars Tertia CCLI–CCCXC*. Edited by Lieven Van Acker and Monika Klaes-Hachmöller. *CC CM* 91B. Turnhout: Brepols, 2001.

———. *Liber divinorum operum*. Edited by Albert Derolez and Peter Dronke. *CC CM* 92. Turnhout: Brepols, 1996.

———. *Liber vite meritorum*. Edited by Angela Carlevaris. *CC CM* 90. Turnhout: Brepols, 1995.

———. *Scivias*. Edited by Adelgundis Führkötter and Angela Carlevaris. *CC CM* 43. Turnhout: Brepols, 1978.

———. *Visiones Hildegardis a Guiberto Gemblacensi exaratae*. In *Hildegardis Bingensis opera minora* II, edited by Jeroen Deploige and Sara Moens, 135–236. *CC CM* 226A. Turnhout: Brepols, 2016.

———. *Vita Sancti Disibodi episcopi*. In *Hildegardis Bingensis opera minora* II, edited by Christopher P. Evans, 59–87. *CC CM* 226A. Turnhout: Brepols, 2016.

———. *Vita Sancti Ruperti confessoris*. In *Hildegardis Bingensis opera minora* II, edited by Christopher P. Evans, 91–108. *CC CM* 226A. Turnhout: Brepols, 2016.

Isidorus Hispalensis. *Etymologiarum libri XX*. In *PL* 82:73–718, edited by Faustinus Arevalus.

Joannes Saresberiensis. *Metalogicon*. Edited by John Barrie Hall and K. S. B. Keats-Rohan. *CC CM* 98. Turnhout: Brepols, 1991.

———. *Metalogicon*. Translated by John Barrie Hall and K.S.B. Keats-Rohan. *CC CM* in translation 12. Turnhout: Brepols, 2013.

———. *The Letters of John of Salisbury*. Edited by William J. Millor, Harold Edgeworth Butler, and Christopher N. L. Brooke. Translated by Christopher N. L. Brooke and Harold Edgeworth Butler. Vol. 1. Oxford, UK: Clarendon Press, 2003 (1955).

Könsgen, Ewald, ed. *Epistolae duorum amantium. Briefe Abaelards und Heloises?* Mittellateinische Studien und Texte 8. Leiden and Cologne: Brill, 1974.

Lefèvre d'Étaples, Jacques, ed. *Liber trium virorum & trium spiritualium virginum*. Paris: ex officina Henrici Stephani, 1513.

Meredith-Jones, Cyril, ed. *Historia Karoli Magni et Rotholandi ou Chronique du Pseudo-Turpin, textes revus et publiés d'après 49 manuscrits*. Paris: Librairie E. Droz, 1936.

Mews, Constant J. *The Lost Love Letters of Heloise and Abelard. Perceptions of Dialogue in Twelfth-Century France*. 2nd ed. Translated by Neville Chiavaroli and Constant J. Mews. The New Middle Ages. New York: Palgrave, 2008 (1999).

Muckle, Joseph T. "Abelard's Letter of Consolation to a Friend (*Historia Calamitatum*)." *Mediaeval Studies* 12 (1950): 163–213.

———. "The Letter of Heloise on Religious Life and Abelard's First Reply." *Mediaeval Studies* 17 (1955): 240–81.

———. "The Personal Letters Between Abelard and Heloise." *Mediaeval Studies* 15 (1953): 47–94.

Mühlbacher, Engelbert, ed. *Die Urkunden Pippins, Karlmanns und Karls des Grossen*. Vol. 1. *MGH*. Hannover: Hahnsche Buchhandlung, 1906.

Nicolaus Claraevallensis. *Epistolae*. In *PL* 196:1593–654, edited by Jean Mabillon.

———. *The Letter Collections of Nicholas of Clairvaux*. Edited and translated by Lena Wahlgren-Smith. Oxford Medieval Texts. Oxford, UK: Oxford University Press, 2018.

Odo de Deogilo. *De profectione Ludovici VII in Orientem: The Journey of Louis VII to the East*. Edited and translated by Virginia Gingerick Berry. Records of Civilization, Sources and Studies. New York: Columbia University Press, 1948.

Petrus Abaelardus. *Peter Abelard. Letters IX-XIV*. Edited by Edmé Renno Smits. Groningen: Rijksuniversiteit, 1983.

———. *Sic et Non*. Edited by Blanch Boyer and Richard McKeon. Chicago: University of Chicago Press, 1976–7.

Petrus Abaelardus and Heloisa Argentoliensis. *Problemata Heloisae*. In *PL* 178:677–730, edited by André Duchesne.

———. *The Letter Collection of Peter Abelard and Heloise*. Edited by David E. Luscombe. Translated by David E. Luscombe and Betty Radice. Oxford Medieval Texts. Oxford, UK: Clarendon Press, 2013.

———. *The Letters of Heloise and Abelard: a Translation of their Collected Correspondence and Related Writings*. Edited and translated by Mary Martin McLaughlin and Bonnie Wheeler. The New Middle Ages. New York: Palgrave, 2009.

Petrus Cellensis. *Epistolae*. In *PL* 202:405–635, edited by Ambroise Janvier.

Petrus Venerabilis. *The Letters of Peter the Venerable*. Edited by Giles Constable. Harvard Historical Studies 78. Cambridge, MA: Harvard University Press, 1967.

Piron, Sylvain, ed. *Lettres des deux amants attribuées à Héloïse et Abélard*. Paris: Gallimard, 2005.

Plechl, Helmut, and Werner Bergmann, eds. *Die Tegernseer Briefsammlung des 12. Jahrhunderts*. MGH. Epistolae 2: Die Briefe der Deutschen Kaiserzeit 2. Hannover: Hahnsche, 2002.

Queneau, Raymond. *Exercices de style*. Gallimard. Paris, 1997 (1947).

Roscelinus Compendiensis. "Der Brief Roscelins an Abälard". In *Der Nominalismus in der Frühscholastik*, edited by Josef Reiners, 63–80. *Beiträge zur Geschichte der Philosophie des Mittelalters* 8.5. Münster: Aschendorffschen Buchhandlung, 1910.

Smyser, Hamilton Martin, ed. *The Pseudo-Turpin, edited from Bibliothèque Nationale, fonds latin, MS. 17656 with an Annotated Synopsis*. The Mediaeval Academy of America 30. Cambridge, MA: The Mediaeval Academy of America, 1937.

Sugerius Sancti Dionysii. *Chartes*. In *Suger. Oeuvres*, edited and translated by Françoise Gasparri, 2:156–267. Les classiques de l'histoire de France au Moyen Age 41. Paris: Les Belles Lettres, 2001.

———. *Gesta Sugerii abbatis (De administratione)*. In *Suger. Oeuvres*, edited and translated by Françoise Gasparri, 1:54–155. Les classiques de l'histoire de France au Moyen Age 37. Paris: Les Belles Lettres, 1996.

———. *Histoire de Louis VII*. In *Suger. Oeuvres*, edited and translated by Françoise Gasparri, 1:156–77. Les classiques de l'histoire de France au Moyen Age 37. Paris: Les Belles Lettres, 1996.

———. *Oeuvres complètes de Suger, recueillies, annotées et publiées d'après les manuscrits, pour la Société de l'Histoire de France*. Edited by Albert Lecoy de la Marche. Paris: Jules Renouard, 1867.

———. *The Deeds of Louis the Fat*. Translated by Richard Cusimano and John Moorhead. Washington D.C.: Catholic University of America Press, 1992.

———. *Vie de Louis VI le gros*. 2nd ed. Edited and translated by Henri Waquet. Les classiques de l'histoire de France au Moyen Age 11. Paris, 1964 (1929).

Theodericus Epternacensis. *Chronicon Epternacense*. In *MGH Scriptores* 23, edited by Ludwig Weiland, 38–64. Hannover, 1874.

———. *Libellus de libertate Epternacensi propugnata*. In *MGH Scriptores* 23, edited by Ludwig Weiland, 64–72. Hannover, 1874.

———. *Octo lectiones in festo sancte Hildegardis legende*. In *Vita sanctae Hildegardis virginis*, edited by Monika Klaes, 75–80. *CC CM* 126. Turnhout: Brepols, 1993.

Titus Lucretius Carus. *De rerum natura libri sex*. Edited by Karl Lachmann. Berlin: George Reimer, 1850.

Waddell, Chrysogonus, ed. *Hymn Collections from the Paraclete*. Cistercian Liturgy Series, 8–9. Gethsemani Abbey, Trappist, KY: Cistercian Publications, 1989.

# Secondary sources


Adler, Alfred. "The *Pelèrinage de Charlemagne* in New Light on Saint-Denis." *Speculum* 22, no. 4 (1947): 550–61.

Allison, Sarah, Ryan Heuser, Matthew Jockers, Franco Moretti, and Michael Witmore. "Quantitative Formalism." In *Canon / Archive*, 1–31. Studies in Quantitative Formalism from the Stanford Literary Lab. New York: n+1 Foundation, 2017.

Andrews, Tara L., and Caroline Macé. "Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation through Graph Analysis of Texts and Stemmata." *Literary and Linguistic Computing* 28, no. 4 (2013): 504–521.

Argamon, Shlomo. "Interpreting Burrows's Delta: Geometric and Probabilistic Foundations." *Literary and Linguistic Computing* 23, no. 2 (2008): 131–47.

Arlot, Sylvain, and Alain Celisse. "A Survey of Cross-Validation Procedures for Model Selection." *Statistics Surveys* 4 (2010): 40–79.

Aschbach, Joseph. *Roswitha und Conrad Celtes*. Vienna: K.K. Hof- und Staatsdruckerei, 1867.

Auerbach, Erich. *Mimesis: Dargestellte Wirklichkeit in der abendländischen Literatur*. 1st ed. Bern: A. Francke A. G., 1946.

———. *Mimesis: The Representation of Reality in Western Literature*. 50th ed. Translated by Willard R. Trask. Princeton: Princeton University Press, 2003 (1946).

Barroux, Robert. "L'Abbé Suger et la vassalité du Vexin en 1124. La levée de l'oriflamme, la Chronique du *pseudo-Turpin* et la fausse donation de Charlemagne à Saint-Denis de 813." *Le Moyen âge: bulletin mensuel d'histoire et de philologie* 64 (1958): 1–26.

Barrow, Julia, Charles S.F. Burnett, and David E. Luscombe. "A Checklist of the Manuscripts Containing the Writings of Peter Abelard and Other Works Closely Associated with Abelard and his School." *Revue d'histoire des textes* 14–5 (1984–5): 183–302.

Barthes, Roland. "The Death of the Author." *Aspen* 5–6 (1967).

Beaulieu, Marie-Anne Polo de. "Colloque *Consensus ex machina*?" *Histoire & Mesure* 9, nos. 1–2 (1994): 171–3.

Benedict, Kimberley M. *Empowering Collaborations: Writing Partnerships between Religious Women and Scribes in the Middle Ages*. Edited by Francis G. Gentry. Studies in Medieval History and Culture 27. New York: Routledge, 2004.

Benton, John F. "A Reconsideration of the Authenticity of the Correspondence of Abelard and Heloise." In *Petrus Abaelardus (1079–1142). Person, Werk und Wirkung*, edited by Rudolf Thomas, 41–52. Trierer Theologische Studien, 38. Trier: Paulinus Verlag, 1980.

———. "Fraud, Fiction and Borrowing in the Correspondence of Abelard and Heloise." In *Pierre Abelard—Pierre le Vénérable: les courants philosophiques, littéraires et artistiques en Occident au milieu du XIIe siècle*, 467–512. Colloques internationaux du Centre national de la recherche scientifique 546. Paris: Éditions du Centre National de la Recherche Scientifique, 1975.

———. "Introduction: Suger's Life and Personality." In *Abbot Suger and Saint-Denis. A Symposium*, edited by Paula Lieber Gerson, 3–15. New York: The Metropolitan Museum of Art, 1986.

———. "Nicolas de Clairvaux." In *Dictionnaire de Spiritualité*, 11:256–59. Paris: Beauchesne, 1982.

———. "Nicolas of Clairvaux and the Twelfth-Century Sequence with Special Reference to Adam of St. Victor." *Traditio* 18 (1962): 149–79.

———. "The Paraclete and the Council of Rouen of 1231." *Bulletin of Medieval Canon Law* 4 (1974): 33–8.

Benton, John F., and Fiorella Prosperetti Ercoli. "The Style of the "Historia Calamitatum": A Preliminary Test of the Authenticity of the Correspondence Attributed to Abelard and Heloise." *Viator* 6 (1975): 59–86.

Berry, David M. "The Computational Turn: Thinking About the Digital Humanities." *Culture Machine* 12 (2011): 1–22.

Bestul, Thomas H. *Texts of the Passion: Latin Devotional Literature and Medieval Society*. Philadelphia: University of Pennsylvania Press, 1996.

Binongo, José Nilo G., and M. Wilfrid A. Smith. "The Application of Principal Components Analysis to Stylometry." *Literary and Linguistic Computing* 14, no. 4 (1999): 446–66.

Bird, Steven, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. Sebastopol: O'Reilly, 2009.

Bod, Rens. *A New History of the Humanities. The Search for Principles and Patterns from Antiquity to Present*. Oxford, UK: Oxford University Press, 2013.

Boglioni, Pierre. "La mort dans les premières hagiographies latines." In *Le sentiment de la mort au moyen âge*, edited by Claude Sutto, 185–210. Montreal: L'aurore, 1979.

Bordalejo, Barbara. "The Genealogy of Texts: Manuscript Traditions and Textual Traditions." *Digital Scholarship in the Humanities* 31, no. 3 (2016): 563–577.

Bournazel, Eric. "Suger and the Capetians." In *Abbot Suger and Saint-Denis. A Symposium*, edited by Paula Lieber Gerson, 55–72. New York: The Metropolitan Museum of Art, 1986.

Boutemy, André. "Le recueil poétique du manuscrit Additional 24199 du British Museum." *Latomus* 2, no. 1 (1938): 30–52.

Brasington, Bruce C. "Lessons of Love: Bishop Ivo of Chartres as Teacher." In *Teaching and Learning in Northern Europe: 1000–1200*, 128–47. Turnhout: Brepols, 2006.

Bredero, Adriaan Hendrik. "Der Brief des heiligen Bernhard auf dem Sterbebett: eine authentische Fälschung." In *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, 5:201–24. Monumenta Germaniae Historica Schriften 33. Hannover: Hahnsche Buchhandlung, 1988.

Breunig, Markus M., Hans-Peter Kriegel, T. Ng Raymond, and Sander Jörg. "LOF: Identifying Density-Based Local Outliers." *ACM SIGMOD Record* 29, no. 2 (2000): 93–104.

Brooke, Christopher N. L. "Approaches to Medieval Forgery." *Journal of the Society of Archivists* 3, no. 8 (1968): 377–86.

Brown, Elizabeth A.R. "*Falsitas pia sive reprehensibilis*. Medieval Forgers and their Intentions." In *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, 1:101–19. Monumenta Germaniae Historica Schriften 33. Hannover: Hahnsche Buchhandlung, 1988.

———. "Saint-Denis and the Turpin Legend." In *The Codex Calixtinus and the Shrine of St. James*, edited by John Williams and Alison Stones, 51–88. Jakobus-Studien 3. Tübingen: Gunter Narr Verlag, 1992.

Brown, Phyllis R., and John C. Peiffer II. "Heloise, Dialectic, and the *Heroides*." In *Listening To Heloise. The Voice of a Twelfth-Century Woman*, edited by Bonnie Wheeler, 143–60. The New Middle Ages. New York: Palgrave, 2000.

Brunn, Uwe. *Des contestataires aux «cathares». Discours de réforme et propagande antihérétique dans les pays du Rhin et de la Meuse avant l'Inquisition*. Collection des études augustiniennes. Série Moyen-Âge et Temps modernes 41. Paris: Institut d'études augustiniennes, 2006.

Buchner, Max. "Das gefälschte Karlsprivileg für St. Denis B M.$^2$ Nr. 482 und seine Entstehung," Historisches Jahrbuch, 42 (1922): 12–28, 250–65.

Bullough, Vern L., Serge Lusignan, and Thomas H. Ohlgren. "Report: Computers and the Medievalist." *Speculum* 49, no. 2 (1974): 392–402.

Bur, Michel. *Suger. Abbé de Saint-Denis, régent de France*. Paris: Librairie Académique Perrin, 1991.

Burrows, John F. "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17, no. 3 (2002): 267–87.

———. "All the Way Through: Testing for Authorship in Different Frequency Strata." *Literary and Linguistic Computing* 22, no. 1 (2007): 27–47.

———. *Computation Into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press, 1987.

———. "The Englishing of Juvenal: Computational Stylistics and Translated Texts." *Style* 36, no. 4 (2002): 677–750.

Bynum, Caroline Walker. "Did the Twelfth Century Discover the Individual?" *Journal of Ecclesiastical History* 31, no. 1 (1980): 1–17.

Caliskan-Islam, Aylin, Richard Harang, Andrew Liu, Arvind Narayanan, Clare Voss, Fabian Yamaguchi, and Rachel Greenstadt. "De-anonymizing Programmers via Code Stylometry." In *Proceedings of the 24th USENIX Security Symposium*, 255–70. Washington, D.C., August 2015.

Calle-Martín, Javier, and Antonio Miranda-García. "Stylometry and Authorship Attribution: Introduction to the Special Issue." *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 251–8.

Campbell, Mary Marshall. "Sanctity and Identity: The Authentication of the Ursuline Relics and Legal Discourse in Elisabeth von Schönau's *Liber Revelationum*." *The Journal of Medieval Religious Cultures* 38, no. 2 (2012): 159–92.

Cartwright, Jane, ed. *The Cult of St Ursula and the 11,000 Virgins*. Cardiff: University of Wales Press, 2016.

Caviness, Madeline H. "Hildegard as Designer of the Illustrations to Her Works." In *Hildegard of Bingen: The Context of Her Thought and Art*, edited by Charles S.F. Burnett and Peter Dronke, 29–62. Warburg Institute Colloquia 4. London: Warburg Institute, 1998.

Cerquiglini, Bernard. "Éloge de la variante." *Langages* 17, no. 69 (1983): 25–35.

———. *Éloge de la variante: Histoire critique de la philologie*. Paris: Seuil, 1989.

———. *In Praise of the Variant: A Critical History of Philology*. 2nd ed. Translated by Betsy Wing. Parallax: Re-Visions of Culture and Society. Baltimore: The John Hopkins University Press, 1999 (1989).

Chandrashekar, Girish, and Ferat Sahin. "A Survey on Feature Selection Methods." *Computers and Electrical Engineering* 40 (2014): 16–28.

Charrier, Charlotte. *Héloise dans l 'histoire et dans la légende*. Bibliothèque de la Revue de littérature comparée 102. Paris: Champion, 1933.

Chazaud, Martial-Alphonse. *Fragments du cartulaire de la Chapelle-Aude*. Moulins: Imprimerie de C. Desroisiers, 1860.

Cheney, Christopher Robert. *The Study of the Medieval Papal Chancery: The Second Edwards Lecture Delivered within the University of Glasgow on 7th December, 1964*. Edwards Lectures 2. Glasgow: Jackson, 1966.

Chénon, Émile. *Histoire et coutumes du prieuré de la Chapelle-Aude*. Paris: Librairie de la société du recueil Sirey, 1915.

Chopra, Deepti, Nisheeth Joshi, and Iti Mathur. *Mastering Natural Language Processing with Python*. Birmingham, UK: Packt Publishing, 2016.

Cioffi, Caron Ann. "The Epistolary Style of Odo of Deuil in his 'De Profectione Ludovici VII in Orientem'." *Mittellateinisches Jahrbuch* 23 (1988): 76–81.

Clanchy, Michael T. *From Memory to Written Record, England 1066–1307*. Oxford, UK: Blackwell, 1993 (1979).

Clark, Anne L. *Elisabeth of Schönau: A Twelfth-Century Visionary*. Middle Ages Series. Philadelphia: University of Pennsylvania Press, 1992.

———. "Repression or Collaboration? The Case of Elisabeth and Ekbert of Schönau." In *Christendom and its Discontents. Exclusion, Persecution, and Rebellion, 1000-1500*, edited by Scott L. Waugh and Peter D. Diehl, 151–67. Cambridge, UK: Cambridge University Press, 1996.

Clausen, Jens Peter. "Suger, faussaire de chartes." In *Suger en question. Regards croisés sur Saint-Denis*, edited by Rolf Große, 109–16. Pariser Historische Studien 68. München: Oldenbourg, 2004.

Clavaud, Florence. "The Digital Edition of the Medieval Charters of the Abbey of Saint-Denis: First Results and Prospects." *Digital Medievalist* 8 (2013). doi:ht tp://doi.org/10.16995/dm.48.

Coakley, John Wayland. "A Shared Endeavour? Guibert of Gembloux on Hildegard of Bingen." In *Women, Men, and Spiritual Power: Female Saints and Their Male Collaborators*, edited by John Wayland Coakley, 45–67. New York: Columbia University Press, 2006.

———. *Women, Men and Spiritual Power: Female Saints and Their Male Collaborators*. New York: Columbia University Press, 2006.

———. "Women's Textual Authority and the Collaboration of Clerics." In *Medieval Holy Women in the Christian Tradition (c. 1100–c.1500)*, edited by Alastair Minnis and Rosalynn Voaden, 83–104. Brepols Essays in European Culture 1. Turnhout: Brepols, 2009.

Coleman, Christopher Bush. *The Treatise of Lorenzo Valla on the Donation of Constantine*. New Haven: Yale University Press, 1922.

Constable, Giles. "Dictators and Diplomats in the Eleventh and Twelfth Centuries: Medieval Epistolography and the Birth of Modern Bureaucracy." *Dumbarton Oaks Papers* 46 (1992): 37–46.

———. "Forged Letters in the Middle Ages." In *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, 5:11–37. Monumenta Germaniae Historica Schriften 33. Primary. Hannover: Hahnsche Buchhandlung, 1988.

———. "Forgery and Plagiarism in the Middle Ages." *Archiv für Diplomatik* 29 (1983): 1–41.

———. *Letters and Letter-Collections*. Typologie des sources du moyen âge occidental 17. Turnhout: Brepols, 1976.

———. "Suger's Monastic Administration." In *Abbot Suger and Saint-Denis. A Symposium*, edited by Paula Lieber Gerson, 17–32. New York: The Metropolitan Museum of Art, 1986.

———. "Sur l'attribution des *Epistolae duorum amantium*." *Comptes rendus des séances de l'Académie des Inscription et Belles-Lettres* 145, no. 4 (2001): 1676–93.

———. "The Authorship of the *Epistolae Duorum Amantium*: A Reconsideration." In *Voices in Dialogue. Reading Women in the Middle Ages*, edited by Linda Steffel Olson and Kathryn Kerby-Fulton, 167–78. Notre Dame, IN: University of Notre Dame Press, 2005.

———. "The Language of Preaching in the Twelfth Century." *Viator* 25 (1994): 131–52.

———, ed. *The Letters of Peter the Venerable*. Harvard Historical Studies 78. London: Harvard University Press, 1967.

———. *The Reformation of the Twelfth Century*. Cambridge, UK: Cambridge University Press, 1996.

Craig, Hugh. "Authorial Attribution and Computational Stylistics: If You Can Tell Authors Apart, Have You Learned Anything About Them?" *Literary and Linguistic Computing* 14, no. 1 (1999): 103–13.

———. "Stylistic Analysis and Authorship Studies." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 273–88. Blackwell Companions to Literature and Culture. Oxford, UK: Blackwell, 2004.

Craig, Hugh, and Arthur F. Kinney. *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge, UK: Cambridge University Press, 2009.

Crane, Gregory. "Classics and the Computer: An End of the History." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 46–55. Blackwell Companions to Literature and Culture. Oxford: Blackwell, 2004.

Crane, Gregory, and Elli Mylonas. "The Perseus Project: An Interactive Curriculum on Classical Greek Civilization." *Educational Technology* 28, no. 11 (1988): 25–32.

Cristianini, Nello, and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press, 2000.

Crown, Ron W. "Comparing the *Patrologia Latina* and the CETEDOC Library of Christian Latin Texts Databases from a User's Perspective." *Journal of Religious & Theological Information* 3, no. 1 (2000): 85–109.

Cunningham, Pádraig, and Sarah Jane Delany. "*k*-Nearest Neighbour Classifiers." *Multiple Classifier Systems* 34 (2007): 1–17.

Cupiccia, Matilde. "Progressi nello studio del cursus: I metodi statistici e il caso di Eloisa e Abelardo." *Filologia mediolatina* 5 (1998): 37–48.

Curtius, Ernst Robert. *European Literature and the Latin Middle Ages*. Translated by Willard R. Trask. Bollingen Series XXXVI. Princeton: Princeton University Press, 2013 (1948).

Daelemans, Walter. "Explanation in Computational Stylometry." In *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, 7817:451–462. Berlin, 2013.

Dalarun, Jacques. "La mort des saints fondateurs. De Martin à François." In *Les fonctions des saints dans le monde occidental (IIIe-XIIIe siècle) Actes du colloque de Rome (27-29 octobre 1988)*, 193–215. Rome: École Française de Rome, 1991.

———. "Nouveaux aperçus sur Abélard, Héloïse et le Paraclet." *Francia* 32, no. 1 (2005): 19–66.

De Gussem, Jeroen. "Bernard of Clairvaux and Nicholas of Montiéramey: Tracing the Secretarial Trail with Computational Stylistics." *Speculum* 92, no. S1 (2017): S190–S225.

Delatte, Louis, and Étienne Evrard. "Un laboratoire d'analyse statistique des langues anciennes à l'Université de Liège." *L'Antiquité Classique* 30–2 (1961): 429–44.

Delatte, Louis, Étienne Evrard, Suzanne Govaerts, and Pierre Hazette. *Sénèque, Consolation à Polybe. Index verborum, relevés statistiques*. The Hague: Mouton, 1962.

Delisle, Léopold. *Les écoles d'Orléans au douzième et au treizième siècle*. L'annuaire-bulletin de la société de l'histoire de France 7. Paris: Imprimerie générale de Ch. Lahure, 1869.

———, ed. *Rouleaux des morts du IXe au XVe siècle*. Société de l'histoire de France. Publications in-8 135. Paris: Mme. Ve. J. Renouard, 1866.

Deploige, Jeroen. "Anonymat et paternité littéraire dans l'hagiographie des Pays-Bas Méridionaux (*ca.* 920 - *ca.* 1320). Autour du discours sur l'«original» et la «copie» hagiographique au Moyen Âge." In *« Scribere sanctorum gesta »: Recueil d'études d'hagiographie médiévale offert à Guy Philippart*, edited by Étienne Renard, Michel Trigalet, Xavier Hermand, and Paul Bertrand, 77–107. Turnhout: Brepols, 2005.

———. "How Gendered Was Clairvoyance in the Thirteenth Century? The Case of Simon of Aulne." In *Speaking to the Eye: Sight and Insight through Text and Image (1150-1650)*, edited by Thérèse de Hemptinne, Veerle Fraeters, and María Eugenia Góngora, 95–126. MISCS 2. Turnhout: Brepols, 2013.

———. *In nomine femineo indocta: Kennisprofiel en ideologie van Hildegard van Bingen (1098-1179)*. Hilversum: Verloren, 1998.

Deploige, Jeroen, Michael Embach, Christopher P. Evans, Kurt Gärtner, and Sara Moens, eds. *Hildegardis Bingensis opera minora* II. *CC CM* 226A. Turnhout: Brepols, 2016.

Derolez, Albert. "Deux notes concernant Hildegarde de Bingen." *Scriptorium* 27 (1973): 291–9.

———. "The Manuscript Transmission of Hildegard of Bingen's Writings: The State of the Problem." In *Hildegard of Bingen: The Context of Her Thought and Art*, edited by Charles S.F. Burnett and Peter Dronke, 17–27. Warburg Institute Colloquia 4. London: Warburg Institute, 1998.

Dessau, Hermann. "Über Zeit und Persönlichkeit der *Scriptores Historiae Augustae*." *Hermes* 24, no. 3 (1889): 337–92.

Di Renzo, Anthony. "His Master's Voice: Tiro and the Rise of the Roman Secretarial Class." *Journal of Technical Writing and Communication* 30, no. 2 (2000): 155–68.

Diederich, Joachim, Jörg Kindermann, Edda Leopold, and Gerhard Paass. "Authorship Attribution with Support Vector Machines." *Applied Intelligence* 19 (2003): 109–23.

Dinzelbacher, Peter. *Vision und Visionsliterature im Mittelalter*. Monographien zur Geschichte des Mittelalters 23. Stuttgart: Hiersemann, 1981.

Doublet, Jacques. *Histoire de l'abbaye de S. Denys en France*. Paris: Chez Jean de Heuqueville, 1625.

Downey, Sarah, Michael D.C. Drout, Veronica E. Kerekes, and Douglas C. Raffle. "Lexomic Analysis of Medieval Latin Texts." *The Journal of Medieval Latin* 24 (2014): 225–275.

Dronke, Peter, ed. *A History of Twelfth-Century Western Philosophy*. Cambridge, UK: Cambridge University Press, 1988.

———. *Abelard and Heloise in Medieval Testimonies*. W. P. Ker Memorial Lecture 26. Glasgow: University of Glasgow Press, 1976.

Dronke, Peter. "Heloise's *Problemata* and *Letters*: Some Questions of Form and Content." In *Petrus Abaelardus (1079–1142). Person, Werk und Wirkung*, edited by Rudolf Thomas, 53–73. Trierer Theologische Studien, 38. Trier: Paulinus Verlag, 1980.

———. *Medieval Latin and the Rise of European Love-Lyric*. 2nd ed. Vol. 2. *Medieval Latin Love-Poetry*. Oxford, UK: Clarendon Press, 1968 (1966).

———, ed. and trans. *Nine Medieval Latin Plays*. Cambridge Medieval Classics 1. Cambridge, UK: Cambridge University Press, 1994.

———. "Peter of Blois and Poetry at the Court of Henry II." *Mediaeval Studies* 38 (1976): 185–235.

———. "Problemata Hildegardiana." *Mittellateinisches Jahrbuch* 16 (1981): 97–131.

———. "The Allegorical World-Picture of Hildegard of Bingen: Revaluations and New Problems." In *Hildegard of Bingen: The Context of Her Thought and Art*, edited by Charles S.F. Burnett and Peter Dronke, 1–16. Warburg Institute Colloquia 4. London: Warburg Institute, 1998.

———. *Women Writers of the Middle Ages: A Critical Study of Texts from Perpetua († 203) to Marguerite Porete († 1310)*. Cambridge, UK: Cambridge University Press, 1984.

———. "Women's Love Letters from Tegernsee." In *Medieval Letters – Between Fiction and Document*, edited by Elisabetta Bartoli and Christian Høgel, 215–45. Utrecht Studies in Medieval Literacy 33. Turnhout: Brepols, 2015.

Dronke, Peter, and Giovanni Orlandi. "New Works by Abelard and Heloise?" *Filologia mediolatina: Studies in Medieval Latin Texts and their Transmission* 12 (2005): 123–77.

Ducey, William Michael. "St. Ailred of Rievaulx and the *Speculum Caritatis*." *The Catholic Historical Review* 17, no. 3 (1931): 308–17.

Dudani, Sahibsingh A. "The Distance-Weighted *k*-Nearest-Neighbor Rule." *IEEE Transactions on Systems, Man, and Cybernetics* 6, no. 4 (1976): 325–7.

Dutton, Paul Edward. *Charlemagne's Mustache and Other Cultural Clusters of a Dark Age*. The New Middle Ages. New York: Palgrave, 2004.

Eco, Umberto. "The Original and the Copy." In *Understanding Origins*, edited by Francisco J. Varela and Jean-Pierre Dupuy. Boston Studies in the Philosophy of Science 130. Dordrecht: Kluwer Academic Publishers, 1992.

Eder, Maciej. "A Bird's Eye View of Early Modern Latin: Distant Reading, Network Analysis and Style Variation'." In *Early Modern Studies after the Digital Turn*, edited by Laura Estill, Diane K. Jakacki, and Michael Ullyot, vol. 502. Medieval and Renaissance Texts and Studies. New Technologies in Medieval and Renaissance Studies 6. Toronto: Iter Press, 2016.

———. "Bootstrapping Delta: a Safety Net in Open-Set Authorship Attribution." In *Digital Humanities 2013: Conference Abstracts*. University of Nebraska–Lincoln, July 2013.

———. "Does Size Matter? Authorship Attribution, Small Samples, Big Problem." *Digital Scholarship in the Humanities* 30, no. 2 (2013): 167–82.

———. "In Search of the Author of the *Chronica Polonorum* Ascribed to Gallus Anonymus: A Stylometric Reconnaissance." *Acta Poloniae Historica* 112 (2015): 5–23.

———. "Mind Your Corpus: Systematic Errors in Authorship Attribution." *Literary and Linguistic Computing* 28, no. 4 (2013): 603–14.

———. "Rolling Stylometry." *Digital Scholarship in the Humanities* 31, no. 3 (2016): 457–69.

———. "Taking Stylometry to the Limits: Benchmark Study on 5,281 Texts from *Patrologia Latina*." In *Digital Humanities 2015: Conference Abstracts*, 1919–1924. Sydney, July 2015.

———. "Visualization in Stylometry: Cluster Analysis Using Networks." *Digital Scholarship in the Humanities* 32, no. 1 (2017): 50–64.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. "Stylometry with R: A Package for Computational Text Analysis." *The R Journal* 8, no. 1 (2016): 107–21.

Elm, Kaspar, and Michel Parisse. *Doppelklöster und andere Formen der Symbiose männlicher und weiblicher Religiosen im Mittelalter*. Berliner historische Studien. Ordensstudien 8. Berlin: Duncker und Humblot, 1992.

Embach, Michael. *Die Schriften Hildegards von Bingen: Studien zu ihrer Überlieferung und Rezeption im Mittelalter und in der Frühen Neuzeit*. Erudiri sapientia: Studien zum Mittelalter und zu seiner Rezeptionsgeschichte 4. Berlin: Akademie Verlag, 2003.

———. "Hildegard of Bingen (1098–1179): A History of Reception." In *A Companion to Hildegard of Bingen*, edited by Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, 273–304. Brill's Companions to the Christian Tradition 45. Leiden: Brill, 2014.

Evans, Gillian Rosemary. *Bernard of Clairvaux*. Great Medieval Thinkers. New York: Oxford University Press, 2000.

Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. "Understanding and Explaining Delta Measures for Authorship Attribution." *Digital Scholarship in the Humanities* 32, no. suppl_2 (2017): ii4–ii16.

Eyheramendy, Susana, David D. Lewis, and David Madigan. "On the Naive Bayes Model for Text Categorization." In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, edited by Christopher M. Bishop and Brendan J. Frey. Key West, January 2003.

Farringdon, Jillian M., Andrew Q. Morton, David M. Baker, and Michael G. Farringdon. *Analysing for Authorship: A Guide to the Cusum Technique*. Cardiff: University of Wales Press, 1996.

Feiss, Hugh. "*Bernardus Scholasticus*: The Correspondence of Bernard of Clairvaux and Hugh of Saint Victor on Baptism." In *Bernardus Magister: Papers Presented at the Nonacentenary Celebration of the Birth of Saint Bernard of Clairvaux, Kalamazoo, Michigan; sponsored by the Institute of Cistercian Studies, Western Michigan Univ., 10-13 May 1990*, 349–78. Cistercian Studies Series 135. Spencer: Cistercian Publications, 1992.

Felten, Franz J. "What Do We Know About the Life of Jutta and Hildegard at Disibodenberg and Rupertsberg?" In *A Companion to Hildegard of Bingen*, edited by Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, 15–38. Brill's Companions to the Christian Tradition 45. Leiden: Brill, 2014.

Ferrante, Joan. "*Scribe quae vides et audis*. Hildegard, Her Language, and Her Secretaries." In *The Tongue of the Fathers: Gender and Ideology in Twelfth-Century Latin*, edited by Townsend David, Taylor Andrew, and Ruth Mazo Karras, 102–35. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 1998.

Ferzoco, George. "The Canonization and Doctorization of Hildegard of Bingen." In *A Companion to Hildegard of Bingen*, edited by Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, 305–16. Brill's Companions to the Christian Tradition 45. Leiden: Brill, 2014.

Fish, Stanley E. "What Is Stylistics and Why Are They Saying Such Terrible Things About It?" In *Approaches to Poetics: Selected Papers from the English Institute*, edited by Seymour Chatman. New York: Columbia University Press, 1973.

———. "What Is Stylistics and Why Are They Saying Such Terrible Things About It?—Part II." *boundary* 2, no. 1 (1979): 129–146.

Flanagan, Sabina. *Hildegard of Bingen, 1098–1179: A Visionary Life*. 2nd ed. London: Routledge, 1998 (1989).

———. "*Spiritualis amicitia* in a Twelfth-Century Convent? Hildegard of Bingen and Richardis of Stade." *Parergon* 29 (1981): 15–21.

Fliche, Augustin. *La réforme grégorienne*. Louvain: Spicilegium sacrum Lovaniense, 1924–37.

Flint, Valerie I.J. "Heinricus of Augsburg and Honorius Augustodunensis: Are They the Same Person?" *Revue bénédictine* 92, nos. 1–2 (1982): 148–58.

———. "The Career of Honorius Augustodunensis: Some Fresh Evidence." *Revue bénédictine* 82, nos. 1–2 (1972): 63–86.

———. "The Chronology of the Works of Honorius Augustodunensis." *Revue bénédictine* 82, nos. 3–4 (1972): 215–42.

———. "The Place and Purpose of the Works of Honorius Augustodunensis." *Revue bénédictine* 87, nos. 1–2 (1977): 97–127.

Flynn, William. "Hildegard (1098–1179) and the Virgin Martyrs of Cologne." In *The Cult of St Ursula and the 11,000 Virgins*, edited by Jane Cartwright, 93–118. Wiltshire: University of Wales Press, 2016.

Folz, Robert. *Le souvenir et la légende de Charlemagne dans l'Empire germanique médiéval*. Publications de l'Université de Dijon 7. Paris: Les Belles Lettres, 1950.

Forstall, Christopher, and Walter Scheirer. *A Statistical Stylistic Study of Latin Elegiac Couplets*, November 2010. `https://www.researchgate.net/publication/251422159`.

Forsyth, Richard S., David I. Holmes, and Emily K. Tse. "Cicero, Sigonio, and Burrows: Investigating the Authenticity of the *Consolatio*." *Literary and Linguistic Computing* 14, no. 3 (1999): 375–400.

Foster, Don. *Author Unknown. On the Trail of Anonymous*. New York: Henry Holt / Company, 2000.

Foucault, Michel. "Qu'est-ce qu'un auteur?" *Bulletin de la société française de philosophie* 69, no. 3 (1969).

Fraeters, Veerle, and Imke de Gier. "Introduction: Shaping Female Spiritual Authority in Europe from the High Middle Ages to the Early Modern Period." In *Mulieres religiosae. Shaping Female Spiritual Authority in the Medieval and Early Modern Periods*, edited by Veerle Fraeters and Imke de Gier, 1–16. Europa Sacra 12. Turnhout: Brepols, 2014.

Fraioli, Deborah A. "George Moore and Scott Moncrieff: An Unknown Chapter in the the Authenticity Debate of the Letters of Abelard and Heloise." *Forum for Modern Language Studies* 54, no. 2 (2017): 176–90.

———. "The Importance of Satire in Jerome's *Adversus Jovinianum* as an Argument against the Authenticity of the *Historia Calamitatum*." In *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, 5:167–200. Monumenta Germaniae Historica Schriften 33. Hannover: Hahnsche Buchhandlung, 1988.

Frischer, Bernard, Donald Guthrie, Emily K. Tse, and Fiona J. Tweedie. "'Sentence' Length and Word-Type at 'Sentence' Beginning and End. Reliable Authorship Discriminators for Latin Prose? New Studies on the Authorship of the *Historia Augusta*." *Research in Humanities Computing* 5 (1996): 110–42.

Funderburg, Kathryn. "Abbot Suger's St. Denis and the Cult of Relics." *The Expositor: A Journal of Undergraduate Research in the Humanities* 4 (2016): 40–7.

Garand, Monique-Cécile. "Auteurs latins et autographes des XI$^{e}$ et XII$^{e}$ siècles." *Scrittura e civiltà* 5 (1981): 77–108.

———. *Guibert de Nogent et ses secrétaires*. Autographa Medii Aeui II. Turnhout: Brepols, 1995.

———. "Le scriptorium de Guibert de Nogent." *Scriptorium* 31, no. 1 (1977): 3–29.

Garay, Kathleen, and Madeleine Jeay. "Sanctification of the Body: The Stages and Staging of Holy Women's Death in High Medieval Europe." In *Heroes and Saints: The Moment of Death in Cross-Cultural Perspectives*, edited by Phyllis Granoff and Koichi Shinohara, 136–58. Newcastle-upon-Tyne: Cambridge Scholars, 2007.

Gasparri, Françoise. "La politique de l'abbé Suger de Saint-Denis à travers ses chartes." *Cahiers de civilisation médiévale* 46, no. 183 (2003): 233–45.

———. "Le latin de Suger, abbé de Saint-Denis (1081–1151)." In *Les historiens et le latin médiéval*, edited by Monique Goullet and Michel Parisse, 177–193. Histoire ancienne et médiévale 63. Paris: Publications de la Sorbonne, 2001.

———, ed. and trans. *Suger. Oeuvres*. Vol. 1. Les classiques de l'histoire de France au Moyen Age 37. Paris: Les Belles Lettres, 1996.

Gatt, Albert, and Emiel Krahmer. "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation." *Journal of Artificial Intelligence Research* 61 (2018): 65–170.

Geary, Patrick J. *Phantoms of Remembrance: Memory and Oblivion at the End of the First Millenium*. Princeton: Princeton University Press, 1994.

George, Philippe. "A Saint-Trond, un import-export de reliques des Onze Mille Vierges dans la seconde moitié du XIIIe siècle." *Bulletin de la Société Royale Le Vieux-Liège* 12, nos. 252-3 (1991): 209–28.

Ghellinck, Joseph de. "Hugues de Saint-Victor." In *Dictionnaire de théologie catholique*, edited by Alfred Vacant and Eugène Mangenot, 7:240–308. Paris: Librairie Letouzey et Ané, 1922.

Gilson, Étienne. *Héloïse et Abélard*. Études sur le moyen age et l'humanisme. Paris: Librairie philosophique J. Vrin, 1938.

Giraud, Cédric, and Constant Mews. "John of Salisbury and the Schools of the 12th Century." In *A Companion to John of Salisbury*, edited by Christophe Grellard and Frédérique Lachaud, 31–62. Brill's Companions to the Christian Tradition 57. Leiden and Boston: Brill, 2015.

Gladwin, Alexander A. G., Matthew J. Lavin, and Daniel M. Look. "Stylometry and Collaborative Authorship: Eddy, Lovecraft, and 'The Loved Dead'." *Digital Scholarship in the Humanities* 29, no. 3 (2014): 443–51.

Glaser, Hubert. "Wilhelm von Saint-Denis: Ein Humanist aus der Umgebung des Abtes Suger und die Krise seiner Abtei von 1151 bis 1153." *Historisches Jahrbuch* 85, no. 2 (1965): 257–322.

Golding, Brian. "Bishops and Nuns: Forms of the *cura monialium* in Twelfth- and Thirteenth-Century England." In *Women in the Medieval Monastic World*, edited by Janet Burton and Karen Stöber, 97–122. Medieval Monastic Studies 1. Turnhout: Brepols, 2015.

Góngora, María Eugenia. "Elisabeth von Schönau and the Story of St Ursula: Visionary Authority and the Cult of the Saints." In *Mulieres religiosae. Shaping Female Spiritual Authority in the Medieval and Early Modern Periods*, edited by Veerle Fraeters and Imke de Gier, 17–35. Europa Sacra 12. Turnhout: Brepols, 2014.

Gössmann, Elisabeth. "'Antiqui' und 'moderni' im 12. Jahrhundert." In *Antiqui und moderni. Traditionsbewußtsein und Fortschrittsbewußtein im späten Mittelalter*, edited by Albert Zimmermann, 40–57. Miscellanea mediaevalia 9. Berlin: De Gruyter, 1973.

———. "Das Menschenbild der Hildegard von Bingen und Elisabeth von Schönau von dem Hintergrund der frühscholastischen Anthropologie." In *Frauenmystik im Mittelalter*, edited by Peter Dinzelbacher and Dieter R. Bauer, 24–47. Stuttgart: Schwabenverlag, 1985.

———. "*Ipsa enim quasi domus sapientiae*: The Philosophical Anthropology of Hildegard von Bingen." *Mystics Quarterly* 13, no. 3 (1987): 146–54.

Goy, Rudolf. *Die Überlieferung der Werke Hugos von St. Viktor: Ein Beitrag zur Kommunikationsgeschichte des Mittelalters*. Monographien zur Geschichte des Mittelalters 14. Stuttgart: Hiersemann, 1976.

Grafton, Anthony. *Forgers and Critics: Creativity and Duplicity in Western Scholarship*. London: Collins & Brown, 1990.

Grant, Lindy. *Abbot Suger of St-Denis: Church and State in Early Twelfth-Century France*. Edited by David Bates. The Medieval World. New York: Addison Wesley Longman, 1998.

Grauert, Hermann. "Novitätenschau." *Historisches Jahrbuch* 12 (1891): 149–236.

Greene, Virginie. "What Happened to Medievalists After the Death of the Author?" In *The Medieval Author in Medieval French Literature*, edited by Virginie Greene, 205–27. Studies in Arthurian and Courtly Cultures. New York: Palgrave, 2006.

Greenspan, Kate. "Autohagiography and Medieval Women's Spiritual Autobiography." In *Gender and Text in the Later Middle Ages*, edited by Jane Chance, 216–36. Gainesville: University Press of Florida, 1996.

Griffin, Robert J. "Anonymity and Authorship." *New Literary History* 30, no. 4 (1999): 877–95.

Griffiths, Fiona J. "'Men's Duty to Provide for Women's Needs': Abelard, Heloise, and their Negotiation of the *cura monialium*." *Journal of Medieval History* 30, no. 1 (2004): 1–24.

———. "Monks and Nuns at Rupertsberg: Guibert of Gembloux and Hildegard of Bingen." In *Partners in Spirit: Women, Men, and Religious Life in Germany, 1100-1500*, edited by Fiona J. Griffiths and Julie Hotchin, 145–69. Medieval Women: Texts and Contexts 24. Turnhout: Brepols, 2014.

———. "The Cross and the *Cura monialium*: Robert of Arbrissel, John the Evangelist, and the Pastoral Care of Women in the Age of Reform." *Speculum* 83, no. 2 (2008): 303–330.

Grossel, Marie-Geneviève. "Voir des yeux du coeur, faire la vision parole: quelques remarques sur les visions d'Elisabeth de Schönau." In *Desir n'a repos. Hommage à Danielle Bohler*, edited by Florence Bouchet and Danièle James-Raoul, 367–77. Eidôlon 115. Bordeaux: Presses universitaires de Bordeaux, 2015.

Groten, Manfred. "Die Urkunde Karls des Großen für St.-Denis von 813 (C 286), eine Fälschung Abt Sugers?" *Historisches Jahrbuch* 108 (1988): 1–36.

Gurevič, Aron Iakovlevič, and Katharine Judelson. *The Origins of European Individualism*. Translated by Katharine Judelson. The Making of Europe. Oxford: Blackwell, 1995.

Gurney, Penelope J., and Lyman W. Gurney. "Authorship Attribution of the *Scriptores Historiae Augustae*." *Literary and Linguistic Computing* 13, no. 3 (1998): 119–31.

Hägermann, Dieter. "Die Urkundenfälschungen auf Karl den Großen: Eine Übersicht." In *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, 3:433–43. Monumenta Germaniae Historica Schriften 33. Hannover: Hahnsche Buchhandlung, 1988.

Haines-Eitzen, Kim. *Guardians of Letters. Literacy, Power, and the Transmitters of Early Christian Literature*. New York: Oxford University Press, 2000.

Hamman, Adalbert-Gautier. "Le centenaire de Jacques Paul Migne (1800–1875)." *Sacris Erudiri* 22 (1974): 13–21.

Haskins, Charles Homer. *The Renaissance of the Twelfth Century*. 8th ed. Cambridge, MA: Harvard University Press, 1971 (1927).

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. Stanford: Springer, 2009 (2001).

Haug, Dag Trygve Truslew, and Marius Larsen Jøhndal. "Creating a Parallel Treebank of the Old Indo-European Bible Translations." In *Proceedings of the Second LREC Workshop on Language Technology for Cultural Heritage Data (LaTeCH)*, edited by Caroline Sporleder and Kiril Ribarov, 27–34. Marrakech, Morocco, June 2008.

He, Sheng, Petros Samara, Jan Burgers, and Lambert Schomaker. "Towards Style-Based Dating of Historical Documents." In *Proceedings of 14th International Conference on Frontiers in Handwriting Recognition*. 2014.

Head, Thomas. *Hagiography and the Cult of Saints: The Diocese of Orléans, 800–1200*. Cambridge, UK: Cambridge University Press, 1990.

Heinermann, Theodor. "Zeit und Sinn der Karlsreise." *Zeitschrift für romanische Philologie* 56, no. 4 (1936): 497–562.

Herrmann, J. Berenike, Karina van Dalen-Oskam, and Christof Schöch. "Revisiting Style: a Key Concept in Literary Studies." *Journal of Literary Theory* 9, no. 1 (2015): 25–52.

Herwegen, Hildephonse. "Les collaborateurs de sainte Hildegarde." *Revue Bénédictine* 21, no. 1 (1904): 192–204.

———. "Les collaborateurs de sainte Hildegarde. *(Fin)*." *Revue Bénédictine* 21, no. 1 (1904): 381–403.

———. "Les collaborateurs de sainte Hildegarde. *(Suite)*." *Revue Bénédictine* 21, no. 1 (1904): 302–316.

Hiatt, Alfred. *The Making of Medieval Forgeries. False Documents in Fifteenth-century England*. London: The British Library, 2004.

Hohler, Christopher. "A Note on *Jacobus*." *Journal of the Warburg and Courtauld Institutes* 35 (1972): 31–80.

Hoover, David L. "Delta prime?" *Literary and Linguistic Computing* 19, no. 4 (2004): 477–95.

———. "The Microanalysis of Style Variation." *Digital Scholarship in the Humanities* 32 (2017): ii17–ii30.

Hotchin, Julie. "Female Religious Life and the *Cura Monialium* in Hirsau Monasticism, 1080 to 1150." In *Listen, Daughter. The Speculum Virginum and the Formation of Religious Women in the Middle Ages*, edited by Constant J. Mews, 59–83. The New Middle Ages. New York: Palgrave, 2001.

Hudson, William Henry. "Hrotsvitha of Gandersheim." *The English Historical Review* 3, no. 11 (1888): 431–57.

Hunter, John D. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9, no. 90 (2007): 90–5.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *Plos One* 9, no. 6 (2014): 1–12.

Jaeger, C. Stephen. *Ennobling Love: In Search of a Lost Sensibility*. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 1999.

———. "*Epistolae duorum amantium* and the Ascription to Heloise and Abelard." In *Voices in Dialogue. Reading Women in the Middle Ages*, edited by Linda Steffel Olson and Kathryn Kerby-Fulton, 125–66. Notre Dame, IN: University of Notre Dame Press, 2005.

———. *The Envy of Angels. Cathedral Schools and Social Ideals in Medieval Europe 950–1200*. Middle Ages Series. Philadelphia: University of Philadelphia Press, 1994.

———. "The *Epistolae Duorum Amantium*, Abelard, and Heloise: An Annotated Concordance." *The Journal of Medieval Latin* 24 (2014): 185–224.

———. "The Prologue to the *Historia Calamitatum* and the 'Authenticity Question'." *Euphorion* 74, no. 1 (1980): 1–15.

Jakobson, Roman, ed. *Structure of Language and Its Mathematical Aspects*. Vol. 12. Proceedings of Symposia in Applied Mathematics. New York: American Mathematical Society, 1961.

Jänicke, Stefan, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. "On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges." In *Proceedings of the Eurographics Conference on Visualization (EuroVis)—STARs*, edited by Rita Borgo, Fabio Ganovelli, and Ivan Viola, 83–103. Cagliari: The Eurographics Association, May 2015.

Janson, Tore. *Prose Rhythm in Medieval Latin from the 9th to the 13th Century*. Acta Universitatis Stockholmiensis. Studia Latina Stockholmiensia 20. Stockholm: Almquist & Wiksell International, 1975.

———. "Schools of Cursus in the Twelfth Century and the Letters of Heloise and Abelard." In *Retorica e poetica tra i secoli XII e XIV. Atti del secondo Convegno internazionale di studi dell'Associazione per il Medioevo e l'Umanesimo latini (AMUL) in onore e memoria di Ezio Franceschini, Trento e Rovereto 3–5 ottobre 1985*, edited by Claudio Leonardi and Enrico Menestò, 171–200. Quaderni del Centro per il Collegamento degli Studi Medioevali e Umanistici nell'Universita di Perugia 18. Perugia, Firenze: «La Nuova Italia» Editrice, 1988.

———. "The Problems of Measuring Sentence-Length in Classical Texts." *Studia Linguistica* 18 (1964): 26–36.

Jeauneau, Edouard. "'Nani gigantum humeris insidentes'. Essai d'interprétation de Bernard de Chartres." *Vivarium* 1 (1967): 79–99.

Jiang, Xia, and Richard E. Neapolitan. *Artificial Intelligence: With an Introduction to Machine Learning*. 2nd ed. Boca Raton: CRC Press, 2018.

Jockers, Matthew L., and Daniela M. Witten. "A Comparative Study of Machine Learning Methods for Authorship Attribution." *Literary and Linguistic Computing* 25, no. 2 (2010): 215–23.

Jockers, Matthew Lee. *Macroanalysis: Digital Methods and Literary History*. Topics in the Digital Humanities. Champaign, IL: University of Illinois Press, 2013.

Johnson, Lynn Staley. "The Trope of the Scribe and the Question of Literary Authority in the Works of Julian of Norwich and Margery Kempe." *Speculum* 66, no. 4 (1991): 820–38.

Juola, Patrick. "Authorship Attribution." *Foundations and Trends in Information Retrieval* 1, no. 3 (2006): 233–334.

Jurafsky, Dan, and James H. Martin. *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd ed. Prentice Hall Series in Artificial Intelligence. Upper Saddle River, New Jersey: Pearson Education International, 2009 (2000).

Karaliutė, Edita. "A Case of Bidirectional Metaphor: A Computer as a Human Being and the Reverse." *Jaunųjų Mokslininkų Darbai* 2, no. 23 (2009): 44–50.

Karian, Stephen. "Authors of the Mind: Some Notes on the QSum Attribution Theory." *Studies in Bibliography* 57 (2005): 263–86.

Kemper, Joachim. "Das benediktinische Doppelkloster Schönau und die Visionen Elisabeths von Schönau." *Archiv für mittelrheinische Kirchengeschichte* 54, no. 1 (2002): 55–102.

Kestemont, Mike. "Function Words in Authorship Attribution: From Black Magic to Theory?" In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL)*, 59–66. Gothenburg, Sweden, April 2014.

———. *Het gewicht van de auteur. Stylometrische auteursherkenning in Middelnederlandse literatuur*. Studies op het gebied van de oudere Nederlandse letterkunde 5. Gent: Koninklijke academie voor Nederlandse taal- en letterkunde, 2013.

———. "Stylometry for Medieval Authorship Studies: An Application to Rhyme Words." *Digital Philology: A Journal of Medieval Cultures* 1, no. 1 (2012): 42–72.

Kestemont, Mike, and Karina van Dalen-Oskam. "Predicting the Past: Memory Based Copyist and Author Discrimination in Medieval Epics." In *Proceedings of the Twenty-First Benelux Conference on Artificial Intelligence*. Edited by Toon Calders, Karl Tuyls, and Mykola Pechenizkiy, 121–8. Eindhoven: BNVKI-AIABN, 2009.

Kestemont, Mike, and Jeroen De Gussem. "Integrated Sequence Tagging for Medieval Latin Using Deep Representation Learning." Edited by Marco Büchler and Laurence Mellerin. *Journal of Data Mining and Digital Humanities*, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages, 2017, 1–17. https://arxiv.org/abs/1603.01597.

Kestemont, Mike, Kim Luyckx, and Walter Daelemans. "Intrinsic Plagiarism Detection Using Character Trigram Distance Scores under a Novel Document Representation." In *Notebook Papers of the 5th Evaluation Lab on Uncovering Plagiarism, Authorship and Social Software Misuse (PAN)*, s.p. Amsterdam, 2011.

Kestemont, Mike, Kim Luyckx, Walter Daelemans, and Thomas Crombez. "Cross-Genre Authorship Verification Using Unmasking." *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 340–356.

Kestemont, Mike, Sara Moens, and Jeroen Deploige. "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux." *Digital Scholarship in the Humanities* 30, no. 2 (2013): 199–224.

Kestemont, Mike, Justin A. Stover, Moshe Koppel, Folgert Karsdorp, and Walter Daelemans. "Authenticating the Writings of Julius Caesar." *Expert Systems with Applications* 63 (2016): 86–96.

Kienzle, Beverly Mayne. "Crisis and Charismatic Authority in Hildegard of Bingen's Preaching against the Cathars." In *Charisma and Religious Authority: Jewish, Christian, and Muslim Preaching, 1200-1500*, edited by Katherine L. Jansen and Miri Rubin, 73–91. Europa Sacra 4. Turnhout: Brepols, 2010.

———. "Defending the Lord's Vineyard: Hildegard of Bingen's Preaching against the Cathars." In *Medieval Monastic Preaching*, edited by Carolyn Muessig, 163–82. Brill's Studies in Intellectual History 90. Leiden: Brill, 1998.

Kleine, Uta. "«Ce sont les mots que profère une langue nouvelle». Élisabeth de Schönau et le renouveau de la prophétie du XIIe siècle." In *Hagiographie et prophétie (VIe-XIIIe siècles)*, edited by Patrick Henriet, Klaus Herbers, and Hans-Christian Lehner, 145–192. Micrologus Library 80. Firenze: Sismel, 2017.

Kohavi, Ron. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 2:1137–43. Montreal, 1995.

Kokhar, Devangana. *Gephi Cookbook*. Birmingham, UK: Packt Publishing, 2015.

Koppel, Moshe, and Jonathan Schler. "Authorship Verification as a One-Class Classification Problem." In *Proceedings of the 21st International Conference on Machine Learning*, s.p. New York: ACM Press, July 2004.

Koppel, Moshe, Jonathan Schler, and Shlomo Argamon. "Computational Methods in Authorship Attribution." *Journal of the American Society for Information Science and Technology* 60, no. 1 (2009): 9–26.

Koppel, Moshe, Jonathan Schler, Shlomo Argamon, and Yaron Winter. "The "Fundamental Problem" of Authorship Attribution." *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 284–91.

———. "The "Fundamental Problem" of Authorship Attribution." *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 284–91.

Koppel, Moshe, and Yaron Winter. "Determining If Two Documents Are Written by the Same Author." *Journal of the Association for Information Science and Technology* 65, no. 1 (2014): 178–87.

Köster, Kurt. "Das visionäre Werk Elisabeths von Schönau. Studien zu Entstehung, Überlieferung und Wirkung in der mitteralterlichen Welt." 4 (1952): 79–119.

———. "Elisabeth von Schönau: Werk und Wirkung im Spiegel der mittelalterlichen handschriftlichen Überlieferung." *Archiv für mittelrheinische Kirchengeschichte* 3 (1951).

———. "Elisabeth von Schönau. Leben, Persönlichkeit und visionäres Werk." In *Schönauer Elisabeth Jubiläum 1965. Festschrift anläßlich des achthundertjährigen Todestages der heiligen Elisabeth von Schönau*, 17–46. Limburg: Pallotiner Druckerei, 1965.

Lalanne, Ludovic. "Quelques doutes sur l'authenticité de la correspondance amoureuse d'Héloïse et d'Abailard." *La correspondence littéraire* 1 (1856): 27–33.

Leclercq, Jean. "Deux épîtres de Saint Bernard et de son secrétaire." In *Recueil d'études sur saint Bernard et ses écrits*, 2:313–8. Storia e Letteratura 104. Rome: Edizioni di Storia e Letteratura, 1966.

———. "Les collections de sermons de Nicolas de Clairvaux." In *Recueil d'études sur saint Bernard et ses écrits*, 1:47–82. Storia e Letteratura 92. Rome: Edizioni di Storia e Letteratura, 1962.

———. "Lettres de S. Bernard: histoire ou littérature?" In *Recueil d'études sur saint Bernard et ses écrits*, 4:125–225. Storia e Letteratura 167. Rome: Edizioni di Storia e Letteratura, 1987.

———. "Modern Psychology and the Interpretation of Medieval Texts." *Speculum* 48, no. 3 (1973): 476–90.

———. "Saint Bernard et ses secrétaires." In *Recueil d'études sur saint Bernard et ses écrits*, 1:3–25. Storia e Letteratura 92. Rome: Edizioni di Storia e Letteratura, 1962.

———. "Toward a Sociological Interpretation of the Various Saint Bernards." In *Bernardus Magister: Papers Presented at the Nonacentenary Celebation of the Birth of Saint Bernard of Clairvaux, Kalamazoo, Michigan; sponsored by the Institute of Cistercian Studies, Western Michigan Univ., 10-13 May 1990*, edited by John Robert Sommerfeldt, 19–33. Cistercian Studies Series 135. Spencer: Cistercian Publications, 1992.

Lerer, Seth. "Medieval English Literature and the Idea of the Anthology." *PMLA* 118, no. 5 (2003): 1251–67.

Levillain, Léon. "Essai sur les origines du Lendit." *Revue Historique* 155, no. 2 (1927): 241–76.

Levison, Wilhelm. *Das Werden der Ursula-Legende*. In *Bonner Jahrbücher* 132, 1–164. 1972.

Liebeschütz, Hans. *Das allegorische Weltbild der Heiligen Hildegard von Bingen*. Leipzig: Teubner, 1930.

Lillich, Meredith Parsons. *The Gothic Stained Glass of Reims Cathedral*. Pennsylvania: The Pennsylvania State University Press, 2011.

Loenertz, Raymond. "Le panégyrique de S. Denys l'Aréopagite par S. Michel le Syncelle." *Analecta Bollandiana* 68 (1950): 94–107.

Loersch, Hugo. "Das falsche Diplom Karls des Grossen und Friedrichs I. Privileg für Aachen vom 8. Januar 1166." In *Die Legende Karls des Grossen im 11. und 12. Jahrhundert. Mit einem Anhang über Urkunden Karls des Grossen und Friechichs I. für Aachen von Hugo Loersch*. 149–215. Publikationen der Gesellschaft für Rheinische Geschichtskunde 7. Leipzig: Verlag von Duncker & Humblot, 1890.

Loomis, Laura Hibbard. "The Oriflamme of France and the War-Cry "Monjoie" in the Twelfth Century." In *Studies in Art and Literature for Belle da Costa Greene*, edited by Dorothy Miner, 67–84. New Jersey: Princeton University Press, 1954.

Love, Harold. *Attributing Authorship. An Introduction*. New York: Cambridge University Press, 2002.

Luscombe, David E. "Denis the Pseudo-Areopagite in the Middle Ages from Hilduin to Lorenzo Valla." In *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, 1:133–52. Monumenta Germaniae Historica Schriften 33. Hannover: Hahnsche Buchhandlung, 1988.

Luyckx, Kim, and Walter Daelemans. "Authorship Attribution and Verification with Many Authors and Limited Data." In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, 513–20. Manchester, UK, August 2008.

———. "Using Syntactic Features to Predict Author Personality from Text." In *Digital Humanities 2008: Conference Abstracts*, 146–9. Oulu, Finland: English Philology, June 2008.

Mabillon, Jean. *De re diplomatica libri VI*. 2nd ed. Paris, 1709 (1681).

Manjavacas, Enrique, Jeroen De Gussem, Walter Daelemans, and Mike Kestemont. "Assessing the Stylistic Properties of Neurally Generated Text in Authorship Attribution." In *Proceedings of the Workshop on Stylistic Variation*, 116–25. Copenhagen, Denmark, September 2017.

Manjavacas, Enrique, Brian Long, and Mike Kestemont. "On the Feasibility of Automated Detection of Allusive Text Reuse." In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 104–14. Minneapolis, MN, 2019.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. New York: Cambridge University Press, 2008.

Marenbon, John. "Authenticity Revisited." In *Listening To Heloise. The Voice of a Twelfth-Century Woman*, edited by Bonnie Wheeler, 19–33. The New Middle Ages. New York: Palgrave, 2000.

———. "Lost Love Letters? A Controversy in Retrospect." *International Journal of the Classical Tradition* 15, no. 2 (2008): 267–80.

Marriott, Ian. "The Authorship of the *Historia Augusta*: Two Computer Studies." *The Journal of Roman Studies* 69 (1979): 65–77.

Martin, Janet. "Classicism and Style in Latin Literature." In *Renaissance and Renewal in the Twelfth Century*, edited by Robert L. Benson, Giles Constable, and Carol Dana Lanham, 537–68. Toronto: University of Toronto Press, 1991 (1982).

Mascol, Conrad. "Curves of Pauline and Pseudo-Pauline Style I." *Unitarian Review* 30 (1888): 452–60.

———. "Curves of Pauline and Pseudo-Pauline Style II." *Unitarian Review* 30 (1888): 539–46.

Masters, Bernadette A. "The Distribution, Destruction and Dislocation of Authority in Medieval Literature and Its Modern Derivatives." *Romanic Review* 82, no. 3 (1991): 270–85.

Mateas, Michael, and Nick Montfort. "A Box, Darkly: Obfuscation, Weird Languages, and Code Aesthetics." In *Proceedings of the 6th Annual Digital Arts and Culture Conference*. Copenhagen, Denmark, December 2005.

Mayr-Harting, Henry. "Odo of Deuil, the Second Crusade and the Monastery of Saint-Denis." In *The Culture of Christendom. Essays in Medieval History in Commemoration of Denis L. T. Bethell*, edited by Marc Anthony Meyer, 225–241. London: Bloomsbury Publishing, 1993.

McGillivray, Barbara, and Adam Kilgarriff. "Tools for Historical Corpus Research, and a Corpus of Latin." In *New Methods in Historical Corpora*, edited by Paul Durell, Martin Scheible, Silke Whitt, and Richard J. Bennett, vol. 3. Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache. Tübingen: Narr, 2013.

McGuire, Brian Patrick. "Bernard's Life and Works: A Review." In *A Companion to Bernard of Clairvaux*, edited by Brian Patrick McGuire, 18–61. Brill's Companions to the Christian Tradition 25. Leiden: Brill, 2011.

———. "Introduction: A Companion to Bernard of Clairvaux." In *A Companion to Bernard of Clairvaux*, edited by Brian Patrick McGuire, 1–17. Brill's Companions to the Christian Tradition 25. Leiden: Brill, 2011.

———. "Loyalty and Betrayal in Bernard of Clairvaux." In *Loyalty in the Middle Ages: Ideal and Practice of a Cross-Social Value*, edited by Jörg Sonntag and Coralie Zermatten, 305–26. Turnhout: Brepols, 2015.

Mehler, Alexander, Tim vor der Brück, Rüdiger Gleim, and Tim Geelhaar. "Towards a Network Model of the Coreness of Texts: An Experiment in Classifying Latin Texts using the TTLab Latin Tagger." In *Text Mining. From Ontology Learning to Automated Text Processing Applications*, edited by Chris Biemann and Alexander Mehler, 87–112. Theory and Applications of Natural Language Processing. Springer, 2014.

Meier, Christel. "Autorschaft im 12. Jahrhundert. Persönliche Identität und Rollenkonstrukt." In *Unverwechselbarkeit. Persönliche Identität und Identifikation in der vormodernen Gesellschaft*, edited by Peter von Moos. Norm und Struktur 23. Cologne: Böhlau Verlag, 2004.

Meißner, Burkhard. "Sum enim unus ex curiosis. Computerstudien zum Stil der *Scriptores Historiae Augustae*." *Rivista di cultura classica e medioevale* 34, no. 1 (1992): 47–79.

Meister, Jan Christoph. "Consensus ex Machina? Consensus *qua* Machina!" *Literary and Linguistic Computing* 10, no. 4 (1995): 263–70.

Mendenhall, Thomas Corwin. "A Mechanical Solution of a Literary Problem." *Popular Science Monthly* 60 (1901): 97–105.

———. "The Characteristic Curves of Composition." *Science* 9, no. 214 (1887): 237–49.

Mews, Constant J. "Hildegard of Bingen and the Hirsau Reform in Germany 1080-1180." In *A Companion to Hildegard of Bingen*, edited by Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco, 57–83. Brill's Companions to the Christian Tradition 45. Leiden: Brill, 2014.

———. "Hugh Metel, Heloise, and Peter Abelard: The Letters of an Augustinian Canon and the Challenge of Innovation in Twelfth-Century Lorraine." *Viator* 32 (2001): 59–92.

———. "Negotiating the Boundaries of Gender in Religious Life: Robert of Arbrissel and Hersende, Abelard and Heloise." *Viator* 37 (2006): 113–48.

———. "The Council of Sens (1141): Abelard, Bernard, and the Fear of Social Upheaval." *Speculum* 77, no. 2 (2002): 342–82.

*MGH Scriptores* 23. Hannover, 1874.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. "Distributed Representations of Words and Phrases and their Compositionality." *Advances in Neural Information Processing Systems* 26 (2013): 3111–9.

Minio-Paluello, Lorenzo. "Iacobus Veneticus Grecus: Canonist and Translator of Aristotle." *Traditio* 8 (1952): 265–304.

Minnis, Alastair J. *Medieval Theory of Authorship: Scholastic Literary Attitudes in the Later Middle Ages*. 2nd ed. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 1988 (1984).

Mitchell, Tom M. *Machine Learning*. New York: McGraw-Hill, 1997.

Moens, Sara. "De horizonten van Guibertus van Gembloers (ca. 1124–1214). De wereld van een benedictijns briefschrijver in tijden van een verschuivend religieus landschap." PhD diss., 2014.

Mohrmann, Christine. "Observations sur la langue et le style de saint Bernard." In *SBO 2*, edited by Jean Leclercq, Henri M. Rochais, and Charles Holwell Talbot, ix–xxxiii. Rome: Editiones Cistercienses, 1958.

Molinier, Auguste. "1848. *Historia gloriosi regis Ludovici*." In *Époque féodale, les Capétiens jusqu'en 1180*. 183. Les Sources de l'histoire de France - Des origines aux guerres d'Italie (1494). 2. Paris: A. Picard et fils, 1902.

Moore, Robert I. *The Origins of European Dissent*. Medieval Academy Reprints for Teaching 30. Toronto: University of Toronto Press, 1994 (1977).

———. *The War on Heresy. Faith and Power in Medieval Europe*. London: Profile Books, 2012.

Moretti, Franco. "Conjectures on World Literature." *New Left Review* 1 (2000): 54–68.

———. *Distant Reading*. London: Verso, 2013.

———. *Graphs, Maps, Trees. Abstract Models for Literary History*. New York: Verso, 2005.

———. "Graphs, Maps, Trees. Abstract Models for Literary History—1." *New Left Review* 24 (2003): 67–93.

———. "Network Theory, Plot Analysis." *New Left Review* 68 (2011): 80–102.

———. "The Slaughterhouse of Literature." *Modern Language Quarterly* 61, no. 1 (2000): 207–27.

Morris, Colin. "Individualism in Twelfth-Century Religion. Some Further Reflections." *The Journal of Ecclesiastical History* 31, no. 2 (1980): 195–206.

———. *The Discovery of the Individual, 1050-1200*. Toronto: University of Toronto Press, 1987 (1972).

Mortensen, Lars Boje. "From Vernacular Interviews Into Prose (ca. 600–1200)." In *Oral Art Forms and Their Passage Into Writing*, edited by Else Mundal and Jonas Wellendorf, 53–68. Copenhagen: Museum Tusculanum Press, 2008.

Morton, Andrew Q., and Sidney Michaelson. *The Qsum Plot*. Technical Report CSR-3-90. Edinburgh: University of Edinburgh, 1990.

Mosteller, Frederick, and David L. Wallace. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. 2nd ed. Springer Series in Statistics. New York: Springer Verlag, 1984 (1964).

Murphy, James J. "The Teaching of Latin as a Second Language in the 12th Century." *Historiographia Linguistica* 7, no. 1 (1980): 159–75.

Naus, James. *Constructing Kingship: The Capetian Monarchs of France and the Early Crusades*. Manchester Medieval Studies. Manchester, UK: Manchester University Press, 2016.

Newman, Barbara. *00.01.06, Mews, The Lost Love Letters of Heloise and Abelard*. The Medieval Review, 2000. Accessed October 6, 2019. `https://scholarworks.iu.edu/journals/index.php/tmr/article/view/14861`.

———. "Authority, Authenticity, and the Repression of Heloise." *Journal of Medieval and Renaissance Studies* 22, no. 2 (1992): 121–57.

———. "Flaws in the Golden Bowl: Gender and Spiritual Formation in the Twelfth Century." *Traditio* 45 (1989): 111–46.

———. "Hildegard and Her Hagiographers. The Remaking of Female Sainthood." In *Gendered Voices. Medieval Saints and Their Interpreters*, edited by Catherine M. Mooney, 16–34. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 1999.

———. *Making Love in the Twelfth Century. Letters of Two Lovers in Context*. Translated by Barbara Newman. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 2016.

———. *Sister of Wisdom. St Hildegard's Theology of the Feminine*. Berkeley and Los Angeles: University of California Press, 1987.

———. "St. Hildegard, Doctor of the Church, and the Fate of Feminist Theology." *Spiritus: A Journal of Christian Spirituality* 13, no. 1 (2013): 36–55.

———. "Three-Part Invention: The *Vita S. Hildegardis* and Mystical Hagiography." In *Hildegard of Bingen: The Context of Her Thought and Art*, edited by Charles S.F. Burnett and Peter Dronke, 189–210. Warburg Institute Colloquia 4. London: Warburg Institute, 1998.

Newman, Mark E. J. *Networks. An Introduction.* Oxford, UK: Oxford University Press, 2010.

Nichols, Stephen G. *From Parchment to Cyberspace. Medieval Literature in the Digital Age.* Medieval Interventions. New Light on Traditional Thinking 2. New York: Peter Lang, 2016.

———. "Introduction: Philology in a Manuscript Culture." *Speculum* 65, no. 1 (1990): 1–10.

———. "Why Material Philology? Some Thoughts." *Zeitschrift für deutsche Philologie* 116 (1997): 10–30.

Nitze, William A. "The So-Called Twelfth-Century Renaissance." *Speculum* 23, no. 3 (1948): 464–71.

Njus, Jesse. "The Politics of Mysticism: Elisabeth of Spalbeek in Context." *Church History* 77, no. 2 (2008): 285–317.

Noecker, John, Michael Ryan, and Patrick Juola. "Psychological Profiling Through Textual Analysis." *Literary and Linguistic Computing* 28, no. 3 (2013): 382–7.

Orlandi, Giovanni. "Le statistiche sulle clausole della prosa. Problemi e proposte," Filologia mediolatina, 5 (1998): 1–36.

———. "Metrica e statistica linguistica come strumenti nel metodo attributivo." *Filologia mediolatina* 6-7 (1999): 9–31.

Pandiri, Thalia A. "Autobiography or Autohagiography? Decoding the Subtext in the Visions of Elisabeth of Schönau." In *Women Writing Latin from Roman Antiquity to Early Modern Europe*, edited by Laurie J. Churchill, Phyllis R. Brown, and Jane E. Jeffrey, 2:197–229. Women Writers of the World. New York: Routledge, 2002.

Parkes, Malcolm Beckwith. "Tachygraphy in the Middle Ages: Writing Techniques Employed for *Reportationes* of Lectures and Sermons." In *Scribes, Scripts and Readers: Studies in the Communication, Presentation and Dissemination of Medieval Texts*, 19–33. London: Hambledon Press, 1991.

Passarotti, Marco Carlo. "Development and Perspectives of the Latin Morphological Analyser LEMLAT." In *Digital Technology and Philological Disciplines*, edited by Andrea Bozzi, Laura Cignoni, and Jean-Louis Lebrave, 20–1:397–414. Linguistica computazionale. Pisa: Istituti Editoriali e Poligrafici Internazionali, 2004.

Passarotti, Marco Carlo. "One Hundred Years Ago. In Memory of Father Roberto Busa SJ." In *Proceedings of The Third Workshop on Annotation of Corpora for Research in the Humanities (ACRH-3)*, edited by Francesco Mambrini, Marco Passarotti, and Caroline Sporleder, 15–24. Sofia, Bulgaria, December 2013.

Patt, William D. "The Early 'Ars Dictaminis' as Response to a Changing Society." *Viator* 9 (1978): 133–56.

Pearsall, Derek. "Forging Truth in Medieval England." In *Cultures of Forgery: Making Nations, Making Selves*, edited by Judith Ryan and Alfred Thomas, 3–14. New York and London: Routledge, 2003.

Pennebaker, James W. *The Secret Life of Pronouns. What Our Words Say About Us*. New York: Bloomsbury Press, 2011.

Pepin, Ronald E. "John of Salisbury as a Writer." In *A Companion to John of Salisbury*, edited by Christophe Grellard and Frédérique Lachaud, 147–79. Brill's Companions to the Christian Tradition 57. Leiden: Brill, 2015.

Piantadosi, Steven T. "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions." *Psychon Bull Rev* 21 (2014): 1112–30.

Piron, Sylvain. "Heloise's Literary Self-Fashioning and the *Epistolae duorum amantium*." In *Strategies of Remembrance: From Pindar to Hölderlin*, edited by Lucie Doležalová, 103–62. Newcastle upon Tyne: Cambridge Scholars Publishing, 2009.

Potha, Nektaria, and Efstathios Stamatatos. "An Improved *Impostors* Method for Authorship Verification." In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, edited by Gareth J.F. Jones, Séamus Lawless, Julio Gonzalo, Liadh Kelly, Lorraine Goeuriot, Thomas Mandl, Linda Cappellato, and Nicola Ferro, 138–45. Dublin: Springer International Publishing, 2017.

Powell, Morgan. "*Vox ex Negativo*. Hildegard of Bingen, Rupert of Deutz and Authorial Identity in the Twelfth Century." In *Unverwechselbarkeit. Persönliche Identität und Identifikation in der vormodernen Gesellschaft*, edited by Peter von Moos, 267–95. Norm und Struktur 23. Cologne: Böhlau Verlag, 2004.

Powicke, Frederick Maurice. "Reviewed Work: *The Pseudo-Turpin, edited from Bibliothèque Nationale, Fonds Latin, MS. 17656 by H. M. Smyser.*" *Speculum* 13, no. 3 (1938): 364–66.

Preger, Wilhelm. *Geschichte der deutschen Mystik im Mittelalter*. Vol. 1. Leipzig: Dörffling und Franke, 1874.

Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. Topics in the Digital Humanities. Champaign, IL: University of Illinois Press, 2011.

Rassow, Peter. "Die Kanzlei Bernhards von Clairvaux." *Studien und Mitteilungen zur Geschichte des Benediktiner-Ordens* 34 (1913): 243–79.

Rauschen, Gerhard. *Die Legende Karls des Grossen im 11. und 12. Jahrhundert. Mit einem Anhang über Urkunden Karls des Grossen und Friechichs I. für Aachen von Hugo Loersch*. Publikationen der Gesellschaft für Rheinische Geschichtskunde 7. Leipzig: Verlag von Duncker & Humblot, 1890.

Richards, Ernest Randolph. *The Secretary in the Letters of Paul*. Wissenschaftliche Untersuchungen zum Neuen Testament 2:42. Tübingen: Mohr, 1991.

Rigg, Arthur G. "Orthography and Pronunciation." In *Medieval Latin : An Introduction and Bibliographical Guide*, edited by Frank A.C. Mantello and Arthur G. Rigg, 79–82. Washington, D.C.: The Catholic University of America Press, 1996.

Ristuccia, Nathan J. "The Herwagen Preacher and his Homiliary." *Sacris Erudiri* 52 (2013): 187–234.

Robertson, Durant Waite. *Abelard and Heloise*. London: Millington, 1972.

Robinson, Peter. "The Digital Revolution in Scholarly Editing." In *Ars Edendi Lecture Series, vol. IV*, edited by Barbara Crostini, Gunilla Iversen, and Brian Møller Jensen, 181–207. Studia Latina Stockholmiensia. Stockholm: Stockholm University Press, 2016.

Rochais, Henri M. "Saint Bernard est-il l'auteur des sermons 40, 41 et 42 «de diversis»?" *Revue Bénédictine* 72, nos. 1–2 (1962): 324–45.

Rorem, Paul. *Hugh of Saint Victor*. Great Medieval Thinkers. New York: Oxford University Press, 2009.

Roth, Ferdinand W.E., ed. *Die Visionen der hl. Elisabeth und die Schriften der Aebte Ekbert und Emecho von Schönau*. Brünn: Raigerner Benediktiner Buchdruckerei. Verlag der Studien aus dem Benediktiner- und Cistercienser-Orden, 1884.

Rouse, Richard H., and Mary A. Rouse. "Wax Tablets." *Language and Communication* 9, nos. 2/3 (1989): 175–91.

Rudman, Joseph. "Non-Traditional Authorship Attribution Studies in the *Historia Augusta*." *Literary and Linguistic Computing* 13, no. 3 (1998): 151–7.

———. "Non-Traditional Authorship Attribution Studies of William Shakespeare's Canon: Some Caveats." *Journal of Early Modern Studies* 5 (2016): 307–28.

Rudman, Joseph. "Non-Traditional Authorship Attribution Studies: Ignis fatuus or Rosetta Stone?" *Bibliographical Society of Australia and New Zealand Bulletin* 24, no. 3 (2000): 163–76.

———. "The State of Authorship Attribution Studies: Some Problems and Solutions." *Computers and the Humanities* 31 (1998): 351–65.

———. "The State of Non-Traditional Authorship Attribution Studies—2012: Some Problems and Solutions." *English Studies. A Journal of English Language and Literature* 93, no. 3 (2012): 259–74.

Ruys, Juanita Feros. "Hearing Mediaeval Voices: Heloise and *Carmina Burana* 126." In *The Poetic and Musical Legacy of Heloise and Abelard*, 91–9. Ottawa: Institute of Mediaeval Music, 2003.

Rybicki, Jan, and Maciej Eder. "Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words?" *Literary and Linguistic Computing* 26, no. 3 (2011): 315–21.

Rybicki, Jan, Maciej Eder, and David L. Hoover. "Computational Stylistics and Text Analysis." In *Doing Digital Humanities*, edited by Constance Crompton, Richard J. Lane, and Ray Siemens, 123–44. London and New York: Routledge, 2016.

Sabersky, Dorette. "The Style of Nicholas of Clairvaux's Letters". In *Erudition at God's Service*, edited by John Robert Sommerfeldt, 195–215. Medieval Cistercian History 11. Cistercian Studies Series 98. Kalamazoo: Cistercian Publications Inc., 1987.

Saenger, Paul. "Silent reading: Its Impact on Late Medieval Script and Society." *Viator* 13 (1982): 367–414.

———. *Space Between Words: The Origins of Silent Reading*. Figurae: Reading Medieval Culture. Stanford, CA: Stanford University Press, 1997.

Sanderson, Conrad, and Simon Guenter. "Short Text Authorship Attribution via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation." In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, 482–491. Sydney, July 2006.

Sanford, Eva Matthews. "Honorius, Presbyter and Scholasticus." *Speculum* 23, no. 3 (1948): 397–425.

Sansone, David. "The Computer and the *Historia Augusta*: A Note on Marriot." *The Journal of Roman Studies*, 1990, 174–7.

Saurma-Jeltsch, Lieselotte. "Die Rupertsberger *Scivias*-Handschrift. Überlegungen zu ihrer Entstehung." In *Hildegard von Bingen. Prophetin durch die Zeiten. Zum 900. Geburtstag*, edited by Edeltraud Forster, 340–58. Basel/Freiburg im Breisgau/Vienna, 1997.

Schepers, Kees. "Abelard's Exegesis of the Song of Songs in his Second Letter to Heloise." *The Journal of Medieval Latin* 27 (2017): 107–31.

Schipperges, Heinrich. "Ein unveröffentlichtes Hildegard-Fragment (Codex Berolin. Lat. Qu. 674)." *Sudhoffs Archiv für Geschichte der Medizin und der Naturwissenschaften* 40, no. 1 (1956): 41–77.

Schmeidler, Bernhard. "Der Briefwechsel zwischen Abälard und Heloise eine Fälschung?" *Archiv für Kulturgeschichte* 11 (1914): 1–30.

Schölkopf, Bernhard, and Alexander J. Smola. *Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond.* Adaptive Computation and Machine Learning. Cambridge, MA: The Massachusetts Institute of Technology Press, 2002.

Schrader, Marianna, and Adelgundis Führkötter. *Die Echtheit des Schrifttums der Heiligen Hildegard von Bingen.* Cologne / Graz: Böhlau Verlag, 1956.

Schulenburg, Jane Tibbets. "Women's Monastic Communities, 500–1100: Patterns of Expansion and Decline." *Signs* 14, no. 2 (1989): 261–92.

Sebastiani, Fabrizio. "Machine Learning in Automated Text Categorization." *ACM Computing Surveys* 34, no. 1 (2002): 1–47.

Shahan, Thomas Joseph. "Constantino Cajetan." In *The Catholic Encyclopedia*, edited by Charles G. Herbermann, 3:360–1. New York: Robert Appleton, 1913.

Silvestre, Hubert. "Die Liebesgeschichte zwischen Abaelard und Heloise: der Anteil des Romans." In *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*, 5:121–65. Monumenta Germaniae Historica Schriften 33. Hannover: Hahnsche Buchhandlung, 1988.

Smith, Peter W.H., and W. Aldridge. "Improving Authorship Attribution: Optimizing Burrows' Delta Method." *Journal of Quantitative Linguistics* 18, no. 1 (2011): 63–88.

Sonzogni, Daniel. "Le chartrier de l'abbaye de Saint-Denis en France au haut Moyen Age. Essai de reconstitution." *Pecia* 3 (2003): 9–211.

Southern, Richard William. *Medieval Humanism and Other Studies*. Oxford: Blackwell, 1970.

Southern, Richard William. *Scholastic Humanism and the Unification of Europe*. Vol. 1. Foundations. Cambridge, MA: Blackwell, 1997 (1995).

Spiegel, Gabrielle M. "The Cult of Saint Denis and Capetian Kingship." *Journal of Medieval History* 1, no. 1 (1975): 43–69.

———. "The *Reditus Regni ad Stirpem Karoli Magni*: A New Look." *French Historical Studies* 7, no. 2 (1971): 145–74.

Spinazzè, Linda. "'Cursus in Clausula': an Online Analysis Tool of Latin Prose." In *Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem*, edited by Francesca Tomasi, Roberto Roselli Del Turco, and Anna Maria Tammaro, vol. 10. New York, 2014.

Spitzer, Leo. *Linguistics and Literary History: Essays in Stylistics*. 2nd ed. New Jersey: Princeton University Press, 1967 (1948).

Stamatatos, Efstathios. "A Survey of Modern Authorship Attribution Methods." *Journal of the American Society for Information Science and Technology* 60, no. 3 (2009): 538–556.

———. "Intrinsic Plagiarism Detection Using Character *n*-gram Profiles." In *SEPLN 2009 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 09)*, edited by Benno Stein, Paolo Rosso, Efstathios Stamatatos, Moshe Koppel, and Eneko Agirre, 38–46. Universidad Politécnica de Valencia, September 2009.

Stamatatos, Efstathios, and Moshe Koppel. "Plagiarism and Authorship Analysis: Introduction to the Special Issue." *Language Resources & Evaluation* 45, no. 1 (2011): 1–4.

Stamatatos, Efstathios, Francisco Rangel, Michael Tschuggnall, Benno Stein, Mike Kestemont, Paolo Rosso, and Martin Potthast. "Overview of PAN 2018." In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, edited by Patrice Bellot, Chiraz Trabelsi, Josiane Mothe, Fionn Murtagh, Jian Yun Nie, Laure Soulier, Eric SanJuan, Linda Cappellato, and Nicola Ferro, 267–85. Cham: Springer International Publishing, 2018.

Stella, Francesco. "Analisi informatiche del lessico e individuazione degli autori nelle *Epistolae duorum amantium* (XII secolo)," 1–10. Oxford, UK, September 2006. `http://www.tdtc.unisi.it/digimed/files/EDA-Statistiche%20Lingua.pdf`.

———. "*Epistolae duorum amantium*: nuovi paralleli testuali per gli inserti poetici." *The Journal of Medieval Latin* 18 (2008): 374–97.

———. "Generic Constants and Chronological Variations in Statistical Linguistics on Latin Epistolography." In *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, edited by Tara Andrews and Caroline Macé, 159–79. Lectio 1. Turnhout: Brepols, 2014.

Stirnemann, Patricia, and Dominique Poirel. "Nicolas de Montiéramey, Jean de Salisbury et deux florilèges d'auteurs antiques." *Revue d'histoire des textes* 1 (2006): 173–88.

Stock, Brian. *The Implications of Literacy: Written Language and Models of Interpretation in the Eleventh and Twelfth Centuries*. Princeton: Princeton University Press, 1983.

Stover, Justin A. "Hildegard, the Schools, and Their Critics." In *A Companion to Hildegard of Bingen*, edited by Beverly Mayne Kienzle, Debra L. Stoudt, and George Ferzoco. Brill's Companions to the Christian Tradition 45. Leiden: Brill, 2014.

Stover, Justin A., and Mike Kestemont. "Reassessing the Apuleian Corpus: A Computational Approach to Authenticity." *The Classical Quarterly* 66, no. 2 (2016): 645–72.

———. "The Authorship of the *Historia Augusta*: Two New Computational Studies." *Bulletin of the Institute of Classical Studies of the University of London*, 2016, 140–57.

Stover, Justin A., Yaron Winter, Moshe Koppel, and Mike Kestemont. "Computational Authorship Verification Method Attributes a New Work to a Major 2nd Century African Author." *Journal of the Association for Information Science and Technology* 67, no. 1 (2016): 239–42.

Sze, Vivienne, Yu-Hsin Chen, Tien-Ju Yang, and Joel S. Emer. "Efficient Processing of Deep Neural Networks: A Tutorial and Survey." *Proceedings of the IEEE (Institute of Electrical and Electronics Engineers)* 105, no. 12 (2017): 2295–2329.

Tempestt, Neal, Sundararajan Kalaivani, Fatima Aneez, Yan Yiming, Xiang Yingfei, and Damon Woodard. "Surveying Stylometry Techniques and Applications." *ACM Computing Surveys* 50, no. 6 (2017): 86:1–86:36.

Thibaut, Jirki. "Intermediary Leadership. The Agency of Abbesses in Ottonian Saxony." In *Abbots and Abbesses as a Human Resource in the Ninth- to Twelfth-Century West*, edited by Steven Vanderputten, 41–60. Vita regularis. Ordnungen und Deutungen religiosen Lebens im Mittelalter 74. Münster: Lit, 2018.

Tombeur, Paul. "*Vox Latina*: Belgian Initiatives in Data Processing the Intellectual Language of Europe, A.D. 197–1965." *Computers and the Humanities* 12 (1978): 13–18.

Tout, Thomas Frederick. "Mediaeval Forgers and Forgeries." In *The Collected Papers of Thomas Frederick Tout with a Memoir and Bibliography*, vol. 3. Lectures. Reprint. Manchester, UK: Manchester University Press, 1934 (1920).

Tse, Emily K., Fiona J. Tweedie, and Bernard Frischer. "Unravelling the Purple Thread: Function Word Variability and the *Scriptores Historiae Augustae*." *Literary and Linguistic Computing* 13, no. 3 (1998): 141–9.

Tuccinardi, Enrico. "An Application of a Profile-Based Method for Authorship Verification: Investigating the Authenticity of Pliny the Younger's Letter to Trajan Concerning the Christians." *Digital Scholarship in the Humanities* 32, no. 2 (2017): 435–47.

Turcan-Verkerk, Anne-Marie. "L'Introduction de l'*Ars Dictaminis* en France. Nicolas de Montiéramey, un professionel du dictamen entre 1140 et 1158." In *Le dictamen dans tous ses états. Perspectives de recherche sur la théorie et la pratique de l'ars dictaminis (xi$^{e}$-xv$^{e}$ siècles)*, edited by Benoît Grévin and Anne-Marie Turcan-Verkerk, 63–98. Turnhout: Brepols, 2015.

Tweedie, Fiona J., and R. Harald Baayen. "How Variable May a Constant Be? Measures of Lexical Richness in Perspective." *Computers and the Humanities* 32 (1998): 323–52.

Tweedie, Fiona J., David I. Holmes, and Thomas N. Corns. "The Provenance of *De Doctrina Christiana*, attributed to John Milton: A Statistical Investigation." *Literary and Linguistic Computing* 13, no. 2 (1998): 77–87.

Unsworth, John. "Medievalists as Early Adopters of Information Technology." *Digital Medievalist* 7 (2012).

Utz, Richard. "*Them* Philologists. Philological Practices and Their Discontents from Nietschze to Cerquiglini." In *The Year's Work In Medievalism*, 26:4–12. 2012.

van 't Spijker, Ineke. *Fictions of the Inner Life. Religious Literature and Formation of the Self in Eleventh and Twelfth Centuries*. Disputatio 4. Turnhout: Brepols, 2004.

Van Acker, Lieven. "Der Briefwechsel der heiligen Hildegard von Bingen Vorbemerkungen zu einer kritischen Edition (*Fortsetzung*)." *Revue Bénédictine* 99, nos. 1–2 (1989): 118–54.

———. "Der Briefwechsel der heiligen Hildegard von Bingen. Vorbemerkungen zu einer Kritischen Edition." *Revue Bénédictine* 98, nos. 1–2 (1988): 141–68.

van Dalen-Oskam, Karina. "Authors, Scribes, and Scholars. Detecting Scribal Variation and Editorial Intervention via Authorship Attribution Methods." In *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, edited by Tara Andrew and Caroline Macé, 141–57. Turnhout: Brepols, 2014.

———. "Epistolary Voices. The Case of Elisabeth Wolff and Agatha Deken." *Literary and Linguistic Computing* 29, no. 3 (2014): 443–51.

———. "The Secret Life of Scribes. Exploring Fifteen Manuscripts of Jacob van Maerlant's *Scolastica* (1271)." *Literary and Linguistic Computing* 27, no. 4 (2012): 355–72.

van Dalen-Oskam, Karina, and Joris van Zundert. "Delta for Middle Dutch - Author and Copyist Distinction in Walewein." *Literary and Linguistic Computing* 22, no. 3 (2007): 345–62.

Van de Kieft, Co. "Deux diplômes faux de Charlemagne pour Saint-Denis, du XIIe siècle." *Le Moyen âge* 64, no. 4 (1958): 401–36.

———. *Étude sur le chartrier et la seigneurie du prieuré de la Chapelle-Aude (xi^e^-xiii^e^ siècle)*. Assen: Van Gorcum & Comp. N.V., 1960.

Van den Eynde, Damien P. "En marge des écrits d'Abélard: les *Excerpta ex regulis Paracletensis monasterii*." *Analecta Praemonstratensia* 38 (1962): 70–84.

Van Engen, John. "Authorship, Authority, and Authorization: The Cases of Abbot Bernard of Clairvaux and Abbes Hildegard of Bingen." *Lectio* 4 (2016): 325–362.

———. "Letters and the Public Persona of Hildegard." In *Hildegard von Bingen in ihrem historischen Umfeld. Internationaler wissenschaftlicher Kongreß zum 900jährigen Jubiläum, 13. – 19. September 1998, Bingen am Rhein*, edited by Alfred Haverkamp and Alexander Reverchon, 375–418. Mainz: Verlag Philipp von Zabern, 2000.

———. *Rupert of Deutz*. Publications of the UCLA Center for Medieval and Renaissance Studies 18. Berkeley, CA: University of California Press, 1983.

———. "Rupert of Deutz and William of Saint-Thierry." *Revue Bénédictine* 93, nos. 3–4 (1983): 327–36.

van Halteren, Hans, Harald Baayen, Fiona J. Tweedie, Marco Haverkort, and Anneke Neijt. "New Machine Learning Methods Demonstrate the Existence of a Human Stylome." *Journal of Quantitative Linguistics* 12, no. 1 (2005): 65–77.

Vanderputten, Steven. *Dark Age Nunneries. The Ambiguous Identity of Female Monasticism, 800–1050*. Ithaca and London: Cornell University Press, 2018.

Verbaal, Wim. "Epistolary Voices and the Fiction of History." In *Medieval Letters – Between Fiction and Document*, edited by Christian Høgel and Elisabetta Bartoli, 1–23. Utrecht Studies in Medieval Literacy 33. Turnhout: Brepols, 2015.

———. "How the West was Won by Fiction: The Appearance of Fictional Narrative and Leisurely Reading in Western Literature (11th and 12th century)." In *True Lies Worldwide: Fictionality in Global Contexts*, edited by Anders Cullhed and Lena Rydholm, 189–202. Berlin: De Gruyter, 2014.

———. "Les sermons sur le cantique de saint Bernard: Un chef d'oeuvre achevé?" *Collectanea cisterciensia* 61 (1999): 167–85.

———. "The Council of Sens Reconsidered: Masters, Monks, or Judges?" *Church History* 74, no. 3 (2005): 460–493.

———. "Trapping the Future: Abelard's Multi-Layered Image-Building." In *Rethinking Abelard: A Collection of Critical Essays*, edited by Babette S. Hellemans, 187–212. Brill's Studies in Intellectual History 229. Leiden and Boston: Brill, 2014.

———. "Voicing your Voice: the Fiction of a Life. Early Twelfth-Century Letter Collections and the Case of Bernard of Clairvaux." *Interfaces* 4 (2017): 103–24.

Verdier, Philippe. "La grande croix de l'abbé Suger à Saint-Denis." *Cahiers de civilisation médiévale* 13, no. 49 (1970): 1–31.

———. "What do we know of the Great Cross of Suger in Saint-Denis*." *Gesta* 9, no. 2 (1970): 12–5.

Verkerk, Philippe, and Anne-Marie Turcan-Verkerk. "Un Programme informatique pour l'étude de la prose rimée e rythmée." *Le Médiéviste et l'Ordinateur* 33 (1996): 41–8.

von Moos, Peter. "Cornelia und Heloise." *Latomus* 34, no. 4 (1975): 1024–59.

———. "Die *Epistolae duorum amantium* und die säkulare Religion der Liebe. Methodenkritische Vorüberlegungen zu einem einmaligen Werk mittellateinischer Briefliteratur." *Studi medievali* 44, no. 1 (2003): 1–116.

———. *Mittelalterforschung und Ideologiekritik. Der Gelehrtenstreit um Heloise*. München: Wilhelm Fink Verlag, 1974.

Waddell, Chrysogonus. "*Epithalamica*: An Easter Sequence by Peter Abelard." *The Musical Quarterly* 72, no. 2 (1986): 239–71.

———, ed. *The Paraclete Statutes: Institutiones Nostrae: Troyes, Bibliothèque Municipale, MS 802, ff. 89r–90v*. Cistercian Liturgy Series 20. Trappist, Kentucky: Cistercian Publications, 1987.

Wahlgren-Smith, Lena. "Editing a Medieval Text: The Case of Nicholas of Clairvaux." In *Challenging the Boundaries of Medieval History: The Legacy of Timothy Reuter*, edited by Patricia Skinner, 173–83. Studies in the Early Middle Ages 22. Turnhout: Brepols, 2009.

Waldman, Thomas G. "Abbot Suger and the Nuns of Argenteuil." *Traditio* 41 (1985): 239–72.

———. "Saint-Denis et les premiers Capétiens." In *Religion et culture autour de l'an Mil*, edited by Dominique Iogna-Prat and Jean-Charles Picard, 191–7. Paris: Picard, 1990.

Walpole, Ronald N. "The *Pèlerinage de Charlemagne*: Poem, Legend, and Problem." *Romance Philology* 8, no. 1 (1954): 173–86.

Wattenbach, Wilhelm. *Das Schrifwesen im Mittelalter*. 3rd ed. Leipzig: Verlag von S. Hirzel, 1896 (1871).

Webb, Geoffrey I. "Naïve Bayes." In *Encyclopedia of Machine Learning*, edited by Geoffrey I. Webb and Claude Sammut, 713–4. New York: Springer, 2011.

Weingart, Scott B. "Demystifying Networks, Parts I & II." *Journal of Digital Humanities* 1, no. 1 (2011). `http://journalofdigitalhumanities.org/1-1/demystifying-networks-by-scott-weingart/`.

Wemple, Suzanne Fonay. *Women in Frankish Society. Marriage and the Cloister 500 to 900*. The Middle Ages. Philadelphia: University of Pennsylvania Press, 1985.

Wetherbee, Winthrop. *Platonism and Poetry in the Twelfth Century. The Literary Influence of the School of Chartres*. Princeton: Princeton University Press, 1972.

Wheeler, Bonnie, ed. *Listening To Heloise. The Voice of a Twelfth-Century Woman*. The New Middle Ages. New York: Palgrave, 2000.

Wheeler, Eric S. "Zipf's Law and Why It Works Everywhere." *Glottometrics* 4 (2002): 45–8.

Widdowson, Henry G. *Stylistics and the Teaching of Literature*. Applied Linguistics and Language Study. New York: Routledge, 2013 (1975).

Widmann, S. "Vita Eckeberti." *Neues Archiv* 11 (1886): 447–54.

Williams, Carrington B. "Mendenhall's Studies of Word-Length Distribution in the Works of Shakespeare and Bacon." *Biometrika* 62, no. 1 (1975): 207–212.

Wilmart, André. "Le dialogue apologétique de moine Guillaume, biographe de Suger (*)." *Revue Mabillon* 32 (1942): 80–118.

Witt, Ronald. "Medieval 'Ars Dictaminis' and the Beginnings of Humanism: a New Construction of the Problem." *Renaissance Quarterly* 35, no. 1 (1982): 1–35.

Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining. Practical Machine Learning Tools and Techniques*. 4th ed. The Morgan Kaufmann Series in Data Management Systems. Cambridge, MA: Morgan Kaufmann Publishers, 2017 (2000).

Wong, Tzu-Tsung. "Performance Evaluation of Classification Algorithms by *k*-fold and Leave-One-Out Cross Validation." *Pattern Recognition* 48 (2015): 2839–2846.

Wulstan, David. "Heloise at Argenteuil and the Paraclete." In *The Poetic and Musical Legacy of Heloise and Abelard*, edited by Marc Stewart and David Wulstan, 67–90. Ottawa: Institute of Mediaeval Music, 2003.

———. "*Novi modulaminis melos*: the Music of Heloise and Abelard." *Plainsong and Medieval Music* 11, no. 1 (2002): 1–23.

Ycart, Bernard. "Alberti's Letter Counts." *Literary and Linguistic Computing* 29, no. 2 (2014): 255–65.

Yu, Bei. "An Evaluation of Text Classification Methods for Literary Study." *Literary and Linguistic Computing* 23, no. 3 (2008): 327–43.

Yule, George Udny. "On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship." *Biometrika* 30, nos. 3/4 (1939): 363–390.

———. *The Statistical Study of Literary Vocabulary*. New York: Archon Books, 1968 (1944).

Zedler, Gottfried. *Die Handschriften der Nassauischen Landesbibliothek zu Wiesbaden*. Zentralblatt für Bibliothekswesen 63. Leipzig: Harrassowitz, 1931.

Zhang, Olivia R., Trevor Cohen, and Scott McGill. "Did Gaius Julius Caesar Write *De Bello Hispaniensi*? A Computational Study of Latin Classics." *Human IT* 14, no. 1 (2018): 28–58.

Zhang, Xinhua. "Support Vector Machines." In *Encyclopedia of Machine Learning*, edited by Claude Sammut and Geoffrey I. Webb, 941–5. New York: Springer, 2011.

Zhang, Yin, Rong Jin, and Zhi-Hua Zhou. "Understanding Bag-of-Words Model: A Statistical Framework." *International Journal of Machine Learning and Cybernetics* 1, no. 1 (2010): 43–52.

Ziolkowski, Jan M. "Cultures of Authority in the Long Twelfth Century." *The Journal of English and Germanic Philology* 108, no. 4 (2009): 421–48.

———. "Lost and Not Yet Found: Heloise, Abelard, and the *Epistolae duorum amantium*." *The Journal of Medieval Latin* 14 (2004): 171–202.

Zipf, George Kingsley. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press, Inc., 1949.

———. *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University Press, 1932.

Zumthor, Paul. *Essai de poétique médiévale*. Paris: Éditions du Seuil, 1972.

# Alphabetical Index

## E



## F



## G



## H

## I



## J



## K



## L



## M



## N



## O



## P

## R



## S



## T



## U



## V



## W



## Y



## Z